## Bilaga 1. Databeredning och skapandet av två simuleringspopulationer

För att genomföra utvärderingen använder vi data från Äbin och skapar målvariabler för samtliga rutor i simuleringspopulationerna för inventeringar som sker två år i rad. Vi betraktar tallungskog och skadad tallungskog som målvariabler och skattar totalen av målvariablerna samt kvoten, dvs. andelen skadad tallungskog för ett antal olika designer.

Genom att känna till om det finns ungskog, tallungskog och skadad tallungskog i alla rutor i populationerna blir det lätt att se olika designers egenskaper med avseende på urvalsfel och bias för skattningarna. Resultaten i rapporten baseras på simuleringar, dvs. ett urval dras med samma design och skattas ett stort antal gångerna för att få fram förväntade värden som kan jämföras med de kända värdena i populationerna.

I studien har data från Äbin år 2015-2022 utnyttjats för att skapa simuleringspopulationerna. Undersökningen har haft samma design sedan starten. (Ett slumpmässigt urval av 1 km x 1 km rutor väljs inom ÄFO.) Vi är intresserade av att skatta antalet ungtallar, skadade ungtallar och andelen skadade ungtallar i simuleringspopulationerna för olika undersökningsstrategier. För att göra det utnyttjar vi Äbin:s skattningar av målvariablerna på 1 km x 1 km rutor. I rapporten kallar vi de unika rutorna för objekt.

Nedan beskrivs hur de två simuleringspopulationerna har skapats och vad som skiljer populationerna åt.

### Simuleringspopulation 1: $\boldsymbol{U}_{\text{Ä}\boldsymbol{bin}}$

Genom att lägga ihop Äbin år 2015-2022 får vi totalt 75 104 rutor. Av dessa är det 1535 rutor som har inventerats två år i rad under åren 2015-2022. Simuleringspopulation 1 är ändlig och betecknas med $U_{\text{Ä}bin}$. Den består av $N_{\text{Ä}bin} = 1535$ objekt (unika rutor) som är märkta med $k = 1, 2, \ldots, \ N_{\text{Ä}bin}$.

För att mäta tillståndsskattningar utgår vi från inventeringar av objekten gjorda vid en tidpunkt, inventeringen kan vara genomförd det första året $t_1$ eller det andra året $t_2$. För att mäta förändringsskattningar mellan två närliggande år behöver objekt $k$ inventeras vid tidpunkterna $t_1$ och $t_2$. Notera att tidpunkterna $t_1$ och $t_2$ inte avser samma år för objekten i populationen. Däremot har vi tillgång till variablerna för alla 1535 objekten i $U_{\text{Ä}bin}$ vid båda tidpunkterna $t_1$ och $t_2$.

### Simuleringspopulation 2: $\boldsymbol{U}_{\boldsymbol{4}\text{Ä}\boldsymbol{FO}}$

För simuleringspopulation 2 har vi för de fyra valda ÄFO:na bara observerat en mätning, den senaste för de objekt som har fler än en mätning. Vi betecknar objektets senaste mätning med $t_1$. Genom att studera och utnyttja olika samband i populationen $U_{\text{Ä}bin}$ skapar vi med hjälp av ett antal stokastiska variabler de nödvändiga variablerna på objektsnivå för den andra mätning, $t_2$, för alla objekten i populationen $U_{4\text{Ä}FO}$. Simuleringspopulationen $U_{4\text{Ä}FO}$ består av $N_{4\text{Ä}FO} = 3070$ objekt som är märkta med $k = 1, 2, \ldots, \ N_{4\text{Ä}FO}$.

Skälet till att det inte går att använda $U_{\text{Ä}bin}$ beror är att det är för få objekt i de fyra valda ÄFO:na som har inventerats två närliggande tidpunkter, $t_1$ och $t_2$. Vilket innebär att vi inte kan redovisa några skattningar på ÄFO-nivå om vi skulle använda simuleringspopulation 1.

För att jämföra olika tillståndsskattningar utgår vi från inventeringar av objekten gjorda vid en tidpunkt, lämpligast är det troligen att använda inventeringsdata från det första året, $t_1$. För att mäta förändringsskattningar mellan $t_1$ och $t_2$ kan vi *inte* testa alla designer eftersom *beståndsarean* för $t_2$ inte har beräknats. Syftet med simuleringspopulation 2 att få kunskap om egenskaperna på ÄFO-nivå för de valda områdena.

## Notering

I praktiken finns det många olika sätt att skapa simuleringspopulationerna på. Andra val hade kunnat ge ett något annorlunda resultat, man bör därför betrakta våra resultat som ett utfall bland fler tänkbara. Tanken är dock att alla designer påverkas på ett likvärdigt sätt när upprepade urval dras ur en population i simuleringsstudien. Vi kan därför uttala oss om vilken av de studerade designerna som fungerar bäst för den aktuella populationen. Den första populationen, $U_{\text{Ä}bin}$, kan betraktas som ett tvärsnitt av alla tillgängliga objekt i Sverige förutsatt att det är slumpen som har gjort att just de aktuella objekten har inventerats två år i rad. I den andra populationen, $U_{4\text{Ä}FO}$, tänker vi att de ingående objekten är slumpmässigt valda ur de fyra ÄFO:na, något som bör vara rimligt. Vi har bara en inventering och skapar värdena för $t_2$ utifrån ett antal stokastiska modeller, vilket ger en modellberoende aspekt att ta hänsyn till i analysen av förändringsskattningar. Detta är något att komma ihåg vid en jämförelse av tänkbara designer mellan simuleringspopulationerna.

## Fördelning av olika egenskaper ger olika utfall för objekten i $\boldsymbol{U_{\text{Ä}bin}}$

Vi börjar med att illustrera tre olika typer av förekomster som påverkar skattningen av målvariablerna för ett objekt: finns det *ungskog* enligt Äbin:s klassificering, finns det *ungtallskog* och har ungtallskogen *betesskador*?

Vi betraktar förekomsten av dessa tre egenskaper i form av sekventiella händelser som beskrivs med tre olika indikatorvariabler (dvs. om objektet har eller saknar egenskapen). Det innebär t.ex. att om det finns betesskador så finns det även ungtallskog och ungskog i objektet. Men att det finns ungskog i objektet betyder inte att det också finns ungtallskog eller betesskador och så vidare. (Notera att inventeringen kan förändra det ursprungliga tillståndet på ungskog som är baserat på flygfoto.) Förekomsten av ungtallskog kräver alltså att det finns ungskog i objektet och förekomsten av ungtallskog med betesskador kräver att det finns ungtallskog och ungskog i objektet. I statistiska sammanhang säger vi att det finns ett beroende mellan de olika förekomsterna (något som utnyttjas i skapandet av $t_2$ i den andra populationen $U_{4\text{Ä}FO}$.)

Vi definierar nu indikatorvariablerna för tidpunkten $t$, som kan vara $t_1$ eller $t_2$. Låt $U_k(t)$ beskriva ungskog enligt Äbin:s klassificering för objekt $k \in U_{\text{Äbin}}$ vid tidpunkten $t$

$$U_k(t) = \begin{cases} 1 & \text{objekt } k \text{ har ungskog} \\ 0 & \text{objekt } k \text{ saknar ungskog} \end{cases}$$

Givet att det finns ungskog i objekt $k$ vid tidpunkten $t$, $U_k(t) = 1$, kan vi definiera indikatorvariabeln $X_k(t)$ som beskriver ifall det finns ungtallskog i objekt $k \in U_{\text{Äbin}}$ vid tidpunkten $t$

$$X_k(t) = \begin{cases} 1 & \text{objekt } k \text{ har ungtallskog, givet att } U_k(t) = 1 \\ 0 & \text{objekt } k \text{ saknar ungtallskog, givet att } U_k(t) = 1 \end{cases}$$

Givet att det finns ungskog och ungtallskog i objekt $k$ vid tidpunkten $t$, $U_k(t) = 1 \; och \; X_k(t) = 1$, kan vi definiera indikatorvariabeln $Y_k(t)$ som beskriver ifall det finns betesskador i ungtallskog i objekt $k \in U_{\text{Äbin}}$ vid tidpunkten $t$

$$Y_k(t) = \begin{cases} 1 & \text{objekt } k \text{ har beteskador, givet att } U_k(t) = 1 \text{ och } X_k(t) = 1 \\ 0 & \text{objekt } k \text{ saknar betesskador, givet att } U_k(t) = 1 \text{ och } X_k(t) = 1 \end{cases}$$

Vi beskriver nu hur objekten i $U_{\text{Äbin}}$ fördelar sig på de tre indikatorvariablerna. I tabell B.1 redovisas populationens andelar (samt antal) av objekt som har eller saknar ungskog för tidpunkterna, $t_1$ och $t_2$. Första årets mätning $U_k(t_1)$ är raderna i tabellen och andra årets $U_k(t_2)$ är kolumnerna.

**Tabell B.1** Fördelning av objekten i $U_{\text{Äbin}}$ över de två tidpunkterna som saknar eller har ungskog, andelar samt antal.

| | $U_k(t_2)$ | | |
|---|---|---|---|
| $U_k(t_1)$ | 0 | 1 | Summa |
| 0 | 0,36 (557) | 0,07 (112) | 0,43 (669) |
| 1 | 0,12 (182) | 0,45 (684) | 0,57 (866) |
| Summa | 0,48 (739) | 0,52 (796) | 1,00 (1535) |

Vi ser i tabell B.1 att andelen av objekten som saknar ungskog enligt Äbin:s definition är 0,36. Vi ser också att träden växer in och ut ur Äbin-beståndet mellan de båda tidpunkterna. Förekomsten av ungskog vid $t_2$ är mer sannolik om det också finns ungskog vid $t_1$, vilket verkar rimligt (jämför andelen 0,07 med 0,45).

I tabell B.2 presenteras populationens fördelning av objekt som har eller saknar tallungskog för tidpunkterna, $t_1$ och $t_2$.

**Tabell B.2** Fördelning av objekten i $U_{\text{Ä}bin}$ över de två tidpunkterna som saknar eller har tallungskog, andelar samt antal.

| | $X_k(t_2)$ | | |
|---|---|---|---|
| $X_k(t_1)$ | 0 | 1 | Summa |
| 0 | 0,46 (708) | 0,09 (132) | 0,55 (840) |
| 1 | 0,11 (166) | 0,34 (529) | 0,45 (695) |
| Summa | 0,57 (874) | 0,43 (661) | 1,00 (1535) |

I tabell B.3 presenteras populationens fördelning av objekt som har eller saknar skadad tallungskog för tidpunkterna, $t_1$ och $t_2$.

**Tabell B.3** Fördelning av objekten i $U_{\text{Ä}bin}$ över de två tidpunkterna som saknar eller har skadad tallungskog, andelar samt antal.

| | $Y_k(t_2)$ | | |
|---|---|---|---|
| $Y_k(t_1)$ | 0 | 1 | Summa |
| 0 | 0,57 (872) | 0,11 (166) | 0,68 (1038) |
| 1 | 0,12 (181) | 0,20 (316) | 0,32 (497) |
| Summa | 0,69 (1053) | 0,31 (482) | 1,00 (1535) |

Komplexiteten i tabell B.2 och B.3 har sitt ursprung i att tallungskog och skadad tallungskog är beroende av tidigare utfall för ungskog som är en flygbildstolkning som dessutom i vissa fall korrigeras vid inventeringen. Vi är intresserade av att uppskatta är våra målvariabler för olika designer. Vi ser att andelen ungtallskog för $t_1$ och $t_2$ varierar mellan 0,43 och 0,45. Andelen skadade ungtallar sjunker till 0,31 respektive 0,32 mellan första och andra inventeringen av populationen. Det är en marginell variation mellan de två inventeringsvarven vilket borde ge likvärda förutsättningar för tillståndsskattningarna. Om vi vill skapa en design med syfte att beräkna förändringsskattningar ser vi att andelen objekt med ungtallar vid båda tidpunkterna är 0,34 och andelen skadade ungtallar vid båda tidpunkter i populationen är 0,20. Om vi kan anta att $U_{\text{Ä}bin}$ speglar den verkliga populationen i Sverige så visar det på relativt låga andelar vilket kan påverka effektivitetsvinsten av att utnyttja permanenta rutor i designen.

Vi är intresserade av att beskriva alla tänkbara kombinationer av de tre indikatorvariablerna för $t_1$ och $t_2$ simultant. För att göra det gör vi följande definitioner av indikatorvariablerna för alla objekt $k \in U_{\text{Ä}bin}$ :

$$u_k\big(U_k(t_1), U_k(t_2)\big), \qquad x_k\big(X_k(t_1), X_k(t_2)\big) \qquad \text{och} \qquad y_k\big(Y_k(t_1), Y_k(t_2)\big)$$

För objekt $k$ har $u_k\big(U_k(t_1), U_k(t_2)\big)$ fyra tänkbara utfall

$$u_k\big(U_k(t_1), U_k(t_2)\big) = \begin{cases} u_k(0,0) & \text{ingen ungskog vid } t_1 \text{ eller } t_2 \\ u_k(0,1) & \text{ingen ungskog vid } t_1 \text{ men vid } t_2 \\ u_k(1,0) & \text{ungskog vid } t_1 \text{ men inte vid } t_2 \\ u_k(1,1) & \text{ungskog vid } t_1 \text{ och } t_2 \end{cases}$$

för objekt $k$ har $x_k\big(X_k(t_1), X_k(t_2)\big)$ fyra tänkbara utfall

$$x_k\big(X_k(t_1), X_k(t_2)\big) = \begin{cases} x_k(0,0) & \text{ingen ungtallskog vid } t_1 \text{ eller } t_2 \\ x_k(0,1) & \text{ingen ungtallskog vid } t_1 \text{ men vid } t_2 \\ x_k(1,0) & \text{ungtallskog vid } t_1 \text{ men inte vid } t_2 \\ x_k(1,1) & \text{ungtallskog vid } t_1 \text{ och } t_2 \end{cases}$$

och för objekt $k$ har $y_k\big(Y_k(t_1), Y_k(t_2)\big)$ fyra tänkbara utfall

$$y_k\big(Y_k(t_1), Y_k(t_2)\big) = \begin{cases} y_k(0,0) & \text{ingen skadad ungtallskog vid } t_1 \text{ eller } t_2 \\ y_k(0,1) & \text{ingen skadad ungtallskog vid } t_1 \text{ men vid } t_2 \\ y_k(1,0) & \text{skadad ungtallskog vid } t_1 \text{ men inte vid } t_2 \\ y_k(1,1) & \text{skadad ungtallskog vid } t_1 \text{ och } t_2 \end{cases}$$

Beteckningarna är lite väl detaljerade men underlättar beskrivningen av de möjliga kombinationerna och programmeringen i Bilaga 2. I tabell B.4 redovisas objektens fördelning på de 16 möjliga kombinationerna av indikatorvariablerna för $U_{\text{Ä}bin}$. Kombinationernas numrering i tabellen används i skapandet av $t_2$ i simuleringspopulation 2.

**Tabell B.4** Fördelning av objekten i $U_{Äbin}$ på de tre indikatorvariablerna.

| Kombination | Ungskog | | Tallar | | Skadade tallar | Antal |
|---|---|---|---|---|---|---|
| (1.1.1) | $u_k(0,0)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 557 |
| (1.2.1) | $u_k(0,1)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 27 |
| (1.2.2) | | | $x_k(0,1)$ | $\rightarrow$ | $y_k(0,0)$ | 31 |
| (1.2.3) | | | | | $y_k(0,1)$ | 54 |
| (1.3.1) | $u_k(1,0)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 55 |
| (1.3.2) | | | $x_k(1,0)$ | $\rightarrow$ | $y_k(0,0)$ | 46 |
| (1.3.3) | | | | | $y_k(1,0)$ | 81 |
| (1.4.1) | $u_k(1,1)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 69 |
| (1.4.2) | | | $x_k(0,1)$ | $\rightarrow$ | $y_k(0,0)$ | 12 |
| (1.4.3) | | | | | $y_k(0,1)$ | 35 |
| (1.4.4) | | | $x_k(1,0)$ | $\rightarrow$ | $y_k(0,0)$ | 16 |
| (1.4.5) | | | | | $y_k(1,0)$ | 23 |
| (1.4.6) | | | $x_k(1,1)$ | $\rightarrow$ | $y_k(0,0)$ | 59 |
| (1.4.7) | | | | | $y_k(0,1)$ | 77 |
| (1.4.8) | | | | | $y_k(1,0)$ | 77 |
| (1.4.9) | | | | | $y_k(1,1)$ | 316 |

## Fördelning av olika egenskaper ger olika utfall för objekten i $\boldsymbol{U_{4ÄFO}}$

Skapandet av $t_2$ använder fördelningarna i tabell B.1-4 samt den observerade fördelningen av objekten i $U_{4ÄFO}$ för tidpunkt $t_1$. Vi betraktar de observerade frekvenserna i tabell B.4 som om de kommer från sannolikhetsfördelningar och skapar binära stokastiska variabler som genererar utfall för objekten vid tidpunkt $t_2$ för de tre indikatorvariablerna.

För ungskog $U_k(t_2)$ betingas fördelningen på om $U_k(t_1) = 0$ eller 1. Eftersom sannolikheten är större att det förekommer ungskog vid tidpunkten $t_2$ om det även finns ungskog vid tidpunkten $t_1$. Genom att utnyttja kombinationerna (1.2.1) till (1.2.3) i tabell B.4 kan de förväntade sannolikheterna $p_x$ och $p_y$ beräknas. Vi skapar Bernoulli-fördelade stokastiska variabler att det finns ungtallskog $X_k(t_2) \sim Be(p_x)$, med sannolikheten $p_x = 85/112 \approx 0{,}76$. På motsvarande sätt beräknas den förväntade sannolikheten $p_y = 54/85 \approx 0{,}64$, att det finns betesskadad tallungskog för $Y_k(t_2) \sim Be(p_y)$. Utfallet av de stokastiska variablerna ger värden på indikatorvariablerna för alla objekten i $U_{4ÄFO}$ med kombinationerna (1.2.1) till (1.2.3). På ett liknande sätt beräknas indikatorvariablerna för övriga kombinationer i $U_{4ÄFO}$.

Genom att betrakta tabell B.4 ser vi att det förekommer två situationer: en lite lättare för de kombinationer som bara beror på indikatorvariablernas värden och en mer komplex situation som kräver att vi utnyttjar ett antal fördelningsantaganden och linjära samband för att kunna beräkna antalet ungtallar och skadade ungtallar. De kombinationer som enbart påverkas av utfallet av indikatorvariabler är: (1.1.1), (1.2.1), (1.3.1), (1.3.2), (1,3,3), (1.4.1), (1.4.4) och (1.4.5). För de övriga kombinationerna hänvisas till SAS-koden i Bilaga 2. En kort sammanfattning är att vi har logaritmerat $X_k(t)$ och $Y_k(t)$ variablerna för att få bättre anpassningar till normalfördelningen, dessa transformeras tillbaka till antalet tallar och skadade tall när variablerna har genererats för $t_2$. För att inte få för avvikande värden i fördelningarna som skapar värdena vid $t_2$ har de använda normalfördelningarna trunkerats. För att hitta övre och undre gränser har fördelningarna för lämpliga kombinationer vid $t_1$ utnyttjats. För ett antal kombinationer, har vi att det finns en symmetri mellan $t_1$ och $t_2$. Vi tänker att kombinationer som bara har ungtallskog och skadade ungtallar vid $t_2$ liknar mer de som bara har förekomster vid $t_1$. Genom att utgå ifrån (1.3.2) kan vi modellera (1.2.2), (1.3.3) kan modellera (1.2.3) och på motsvarande sätt kan (1.4.2), (1.4.3) och (1.4.7) modelleras. För kombinationerna (1.4.6) till (1.4.9) utnyttja objektens fördelningar och linjära samband i $U_{Äbin}$ och $U_{4ÄFO}$ för att skapa antalet ungtallar och skadade ungtallar för objekten $t_2$.

I tabell B.5 redovisas objektens fördelning på de 16 möjliga kombinationerna av indikatorvariablerna för $U_{4ÄFO}$.

**Tabell B.5** Fördelning av objekten i $U_{4ÄFO}$ på de tre indikatorvariablerna.

| Kombination | Ungskog | | Tallar | | Skadade tallar | Antal |
|---|---|---|---|---|---|---|
| (1.1.1) | $u_k(0,0)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 1343 |
| (1.2.1) | $u_k(0,1)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 75 |
| (1.2.2) | | | $x_k(0,1)$ | $\rightarrow$ | $y_k(0,0)$ | 81 |
| (1.2.3) | | | | | $y_k(0,1)$ | 146 |
| (1.3.1) | $u_k(1,0)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 64 |
| (1.3.2) | | | $x_k(1,0)$ | $\rightarrow$ | $y_k(0,0)$ | 72 |
| (1.3.3) | | | | | $y_k(1,0)$ | 149 |
| (1.4.1) | $u_k(1,1)$ | $\rightarrow$ | $x_k(0,0)$ | $\rightarrow$ | $y_k(0,0)$ | 167 |
| (1.4.2) | | | $x_k(0,1)$ | $\rightarrow$ | $y_k(0,0)$ | 18 |
| (1.4.3) | | | | | $y_k(0,1)$ | 65 |
| (1.4.4) | | | $x_k(1,0)$ | $\rightarrow$ | $y_k(0,0)$ | 24 |
| (1.4.5) | | | | | $y_k(1,0)$ | 38 |
| (1.4.6) | | | $x_k(1,1)$ | $\rightarrow$ | $y_k(0,0)$ | 76 |
| (1.4.7) | | | | | $y_k(0,1)$ | 181 |
| (1.4.8) | | | | | $y_k(1,0)$ | 144 |
| (1.4.9) | | | | | $y_k(1,1)$ | 427 |

## Bilaga 2. SAS-program för skapandet simuleringspopulation 2

```
/*******************************************************************************************\

              SKOGSSTYRELSEN: ÄBIN uppdrag 2023

        Uttag ur databasen för inventeringsåren 2015-2022 samtliga ÄFO

        Skapandet av simulerad population för fyra ÄFO

        Excelfiler på ruta-nivå från Neil

\*******************************************************************************************/

/*** Spara SAS-dataset i katalog ***/
libname c "C:\Users\...\SAS_DATA";


/** Skapar SASfil för Analysfil Rutor **/
PROC IMPORT OUT= WORK.A_rutor
 DATAFILE= "C:\Users\...\Excel_DATA\ABINANALYSInvRutor.xlsx"  DBMS=EXCEL REPLACE;
    RANGE="Query$";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;

/** Genererad från: ABINANALYSInvRutor.xlsx (75104 obs) **/
data c.ABIN_Analys_rutor;
 set A_rutor;
  Ruta_ID=substr(RutaRefID,3,8);
run;


/***** Testar om två mätningar är gjorda på samma ruta *****/
data match(keep=LottadeRutaID RutaID_old InvAR_old);
 set c.Abin_rutor;
  RutaID_old=RutaID;
  InvAR_old=InvAr;
run;
proc sort data=match out=match1;
 by RutaID_old InvAR_old;
run;
data match2;
 set match1;
 by RutaID_old InvAr_old;
 if first.RutaID_old then count=0;
  count+1;
run;
```

```
/**** sortera för att matcha på RutaID_old ****/
proc sort data=c.ABIN_Analys_rutor out=Ruta_year;
 by LottadeRutaID;
run;
proc sort data=match2;
 by LottadeRutaID;
run;

data Ruta_year_1;
 merge Ruta_year(in=a) match2(in=b);
 by LottadeRutaID;
  if a and b then output;
run;

proc sort data=Ruta_year_1;
 by RutaID_old count;
run;

data test t1 t2(keep=RutaID_old year2 inv_ok2 Ant_Tall_tot2 Ant_Tall_skada2 yk_2 xk_2);
 set Ruta_year_1;
  by RutaID_old count;

  if count=1 then do;
		year1=Invar+0;
		Inv_ok1=(InvKriterieSUM>0);
		Ant_Tall_tot1=AntalTallar_Antal+0;
		Ant_Tall_skada1=TArskada_Antal+0;
		xk_1=Tallar_ViktadRuta+0;
		yk_1=TArskada_ViktadRuta+0;
   output t1;
  end;
  if count=2 then do;
		year2=Invar+0;
		Inv_ok2=(InvKriterieSUM>0);
		Ant_Tall_tot2=AntalTallar_Antal+0;
		Ant_Tall_skada2=TArskada_Antal+0;
		xk_2=Tallar_ViktadRuta+0;
		yk_2=TArskada_ViktadRuta+0;
   output t2;
  end;

  output test;
run;
```

```
/*******************************************************************************************
      Population tidpunkt 1: före genereradet av värden för tidpunkt 2, för de fyra ÄFO
********************************************************************************************/

/* Antal obs: 3070 */
data sim_pop_0 (Keep = RutaID_old AFO DelomrNr SWEREF99Nord SWEREF99Ost BestAreaSUM1 BestKorrAllaSUM1 year1 I_ungskog1
Ant_Tall_tot1 Ant_Tall_skada1 xk_1 ln_x1 I_x1 yk_1 ln_y1 I_y1);
 set test;
  by RutaID_old count;
   AFO="00-00";
   if Lankod="13" and AFONr="3" then AFO="13-03";
   if Lankod="20" and AFONr="16" then AFO="20-16";
   if Lankod="23" and AFONr="2" then AFO="23-02";
   if Lankod="24" and AFONr="5" then AFO="24-05";

            if AFO ne "00-00" then do;
             if last.count then do;
              BestAreaSUM1=BestAreaSUM+0;
               if BestAreaSUM1<0 then BestAreaSUM1=0;*för missing;
              BestKorrAllaSUM1=BestKorrAllaSUM+0;
               if BestKorrAllaSUM1<0 then BestKorrAllaSUM1=0;*för missing;
              year1=Invar+0;
              I_ungskog1=(InvKriterieSUM>0);*ungskog tidpunkt 1;
              Ant_Tall_tot1=AntalTallar_Antal+0;
               if Ant_tall_tot1<0 then Ant_tall_tot1=0;*för missing;
              Ant_Tall_skada1=TArskada_Antal+0;
               if Ant_Tall_skada1<0 then Ant_Tall_skada1=0;*för missing;
              xk_1=Tallar_ViktadRuta+0;
              ln_x1=log(xk_1+1);
              I_x1=(xk_1>0);
              yk_1=TArskada_ViktadRuta+0;
              I_y1=(yk_1>0);
              ln_y1=log(yk_1+1);

              if (yk_1<=xk_1) then output;

             end;
            end;
run;
```

```
/*****************************************************************************************************
        Skapar indikatorvariablerna I_ungskog2, I_x2 och I_y2 för t2 och genererar
      xk_2 och yk_2 för (1.1.1) (1.2.1) (1.3.1) (1.3.2) (1.3.3) (1.4.1) (1.4.4) och (1.4.5)
*****************************************************************************************************/

data sim_pop_1;
 set sim_pop_0;
  call streaminit(123);
  if I_ungskog1=0 then I_ungskog2 = rand("Bernoulli",0.17);
  if I_ungskog1=1 then I_ungskog2 = rand("Bernoulli",0.79);

  xk_2=0;ln_x2=0;
  yk_2=0;ln_y2=0;

  /*1.1*/
  if I_ungskog1=0 and I_ungskog2=0 then do;
    I_x2=0;I_y2=0;
  end;
  /*1.2*/
  if I_ungskog1=0 and I_ungskog2=1 then do;
              I_x2 = rand("Bernoulli",0.76);
               /*1.2.1*/
               if I_x2=0 then do;
                xk_2=0;ln_x2=0;
                I_y2=0;yk_2=0;ln_y2=0;
               end;
               if I_x2=1 then do;
                I_y2 = rand("Bernoulli",0.64);
                 /*1.2.2*/ *Beräknas i c.sim_pop_3070_brutto;
                 /*1.2.3*/ *Beräknas i c.sim_pop_3070_brutto;
               end;/*I_x2=1*/
  end;/*1.2*/

  /*1.3*/
  if I_ungskog1=1 and I_ungskog2=0 then do;
    I_x2=0;xk_2=0;ln_x2=0;
    I_y2=0;yk_2=0;ln_y2=0;
  end;

  /*1.4*/
  if I_ungskog1=1 and I_ungskog2=1 then do;
     if I_x1=0 then do;
               I_x2 = rand("Bernoulli",0.4);
                 /*1.4.1*/
               if I_x2=0 then do;
                 xk_2=0;ln_x2=0;
                 I_y2=0;yk_2=0;ln_y2=0;
               end;
       /*1.4.2-3*/
       if I_x2=1 then do;
         I_y2 = rand("Bernoulli",0.74);
                   /*1.4.2*/  *Beräknas i c.sim_pop_3070_brutto;
               /*1.4.3*/  *Beräknas i c.sim_pop_3070_brutto;
       end;/*I_x2=1*/
     end;/*I_x1=0*/
```

```
      if I_x1=1 then do;
                   I_x2 = rand("Bernoulli",0.93);
                    /*1.4.4-5*/
                   if I_x2=0 then do;
                    xk_2=0;ln_x2=0;
                    I_y2=0;yk_2=0;ln_y2=0;
                   end;
                    /*1.4.6-9*/
       if I_x2=1 then do;
         I_y2 = rand("Bernoulli",0.745);
                            /*1.4.6*/  *Beräknas i c.sim_pop_3070_brutto;
                            /*1.4.7*/  *Beräknas i c.sim_pop_3070_brutto;
                            /*1.4.8*/  *Beräknas i c.sim_pop_3070_brutto;
                            /*1.4.9*/  *Beräknas i c.sim_pop_3070_brutto;
       end;/*I_x2=1*/
      end;/*I_x1=1*/

    end;/*1.4*/
  p_x1x2=11*(I_x1=1)*(I_x2=1)+10*(I_x1=1)*(I_x2=0)+1*(I_x1=0)*(I_x2=1);
  p_y1y2=11*(I_y1=1)*(I_y2=1)+10*(I_y1=1)*(I_y2=0)+1*(I_y1=0)*(I_y2=1);
run;

/*********************************************************************************************
        Fördelningar och samband för att skapa resterande kombinationer
         värden beräknade insatta i c.sim_pop_3070_brutto
*********************************************************************************************/

/**** Fördelningar från (1.3.2) skapar (1.2.2) ****/
title "   /* Fall 1.3.2 */ where((I_ungskog1=1)*(I_ungskog2=0)*(p_x1x2=10)*(p_y1y2=0)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=0)*(p_x1x2=10)*(p_y1y2=0));
  var ln_x1;
  histogram ln_x1/normal (color=red w=4);
run;
/**** Fördelningar från (1.3.3) skapar (1.2.3) ****/
title "   /* Fall 1.3.3 */ where((I_ungskog1=1)*(I_ungskog2=0)*(p_x1x2=10)*(p_y1y2=10)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=0)*(p_x1x2=10)*(p_y1y2=10));
  var ln_x1 ln_y1;
  histogram ln_x1/normal (color=red w=4);
run;
/*  Reg på (1.3.3) till ln_y2 i (1.2.3) */
proc reg data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=0)*(p_x1x2=10)*(p_y1y2=10));
 model ln_y1 = ln_x1/noint; * tar bort inerceptet;
run;
quit;
/**** Fördelningar från (1.4.4) skapar (1.4.2) ****/
title "   /* Fall 1.4.4 */ where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=10)*(p_y1y2=0)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=10)*(p_y1y2=0));
  var ln_x1;
  histogram ln_x1/normal (color=red w=4);
run;
```

```
/**** Fördelningar från (1.4.5) skapar (1.4.3) ****/
title "   /* Fall 1.4.5 */ where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=10)*(p_y1y2=10)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=10)*(p_y1y2=10));
  var ln_x1 ln_y1;
  histogram ln_x1/normal (color=red w=4);
run;
/*  Reg på (1.4.5) till ln_y2 i (1.4.3) */
proc reg data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=10)*(p_y1y2=10));
 model ln_y1 = ln_x1/noint; * tar bort inerceptet;
run;
quit;
/**** Fördelningar från Pop_1535 till beräkningar på t1 för att skatta Regr i (1.4.6) ****/
title "   /* Fall 1.4.6 */ where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=0)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=0));* skattat från förd. i Pop_1535;
  var ln_x1;*Min och Max från Pop_1535;
  histogram ln_x1/normal (color=red w=4);
run;

/**** Fördelningar från Pop_1535 till beräkningar på t1 för att skatta Regr i (1.4.7) ****/
title "   /* Fall 1.4.7 */ where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=1)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=1));
  var ln_x1;
run;
title "   /* Fall 1.4.8 */ where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=10)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=10));
  var ln_x1 ln_y1;
run;
/*  Reg på (1.4.8) till ln_y2 i (1.4.7) */
proc reg data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=10));
 model ln_y1 = ln_x1/noint; * tar bort inerceptet;
run;
quit;
title "   /* Fall 1.4.9 */ where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=11)) **";
proc univariate Normal plot data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=11));
  var ln_x1 ln_y1;
run;
proc reg data=sim_pop_1;
where((I_ungskog1=1)*(I_ungskog2=1)*(p_x1x2=11)*(p_y1y2=11));
 model ln_y1 = ln_x1/noint; * tar bort inerceptet;
run;
quit;
```

```
/****************************** Skapar t2 i Pop_3070 (4-ÄFO) ***********************************/
/*   fixar xk_2 och yk_2 i (1.2.2) (1.2.3) (1.4.2) (1.4.3) (1.4.6) (1.4.7) (1.4.8) och (1.4.9)  */
/*******************************************************************************************/
data c.sim_pop_3070_brutto;
 set sim_pop_1;
  call streaminit(1234);

   y2_ok=1;*flagga för att yk_2<=xk_2;
     /*1.2*/
    if I_ungskog1=0 and I_ungskog2=1 then do;
     /*1.2.2-3*/
     if I_x2=1 then do;
               /*1.2.2*/
               if I_y2=0 then do;
                         a=1*4.65;/*(1.3.2) för min=4.65 */
                         Fa=cdf("Normal",a,7.19,1.36);/* NF från (1.3.2) */
                         b=1*10.7;/*(1.3.2) för max=10.7 */
                         Fb=cdf("Normal",b,7.19,1.36);

                         v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                         ln_x2 = quantile("Normal",v,7.19,1.36); /* trunkerad NF från (1.3.2) på [min,max] */
                         xk_2=exp(ln_x2);
                         yk_2=0;ln_y2=0;
               end;

               /*1.2.3*/
               if I_y2=1 then do;
                        /* Ungtall */
                         a=1*4.83;/*(1.3.3) för min=4.83 */
                         Fa=cdf("Normal",a,8.74,1.35);
                         b=1*11.53;/*(1.3.3) för max=11.53 */
                         Fb=cdf("Normal",b,8.74,1.35);

                         v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                         ln_x2 = quantile("Normal",v,8.74,1.35); /* trunkerad NF från (1.3.3) på [min,max] */
                         xk_2=exp(ln_x2);

       /* Skaddade ungtallar */
*       e_k=rand("Normal",0,1.01);* regression (1.3.3);
*       ln_y2 = 0.7516*ln_x2 + e_k;
                         hat_y2=0.7516*ln_x2;
                         a=1*4.57;/*(1.3.3) för min=4.57*/
                         Fa=cdf("Normal",a,hat_y2,1.01);
                         b=1*9.67;/*(1.3.3) för max=9.67*/
                         Fb=cdf("Normal",b,hat_y2,1.01);

                         v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                         ln_y2 = quantile("Normal",v,hat_y2,1.01); /* trunkerad NF på [min,max] */

                         if ln_y2>ln_x2 then do;
                           y2_ok=0;
                           ln_y2=(0.8 + 0.2*rand("Uniform"))*ln_x2;*kan behövas en justering F(a)=0.8 F(b)=1;
                         end;
                         yk_2=exp(ln_y2);
               end; /*1.2.3*/
     end; /*I_x2=1*/
    end; /*1.2*/

   /*1.4*/
   if I_ungskog1=1 and I_ungskog2=1 then do;
     if I_x1=0 then do;
      /*1.4.2-3*/
      if I_x2=1 then do;
                /*1.4.2*/
                if I_y2=0 then do;
```

```
                              a=1*4.95;/*(1.4.4) för min=4.95 */
                              Fa=cdf("Normal",a,6.86,1.46);
                              b=1*9.17;/*(1.4.4) för max=9.17 */
                              Fb=cdf("Normal",b,6.86,1.46);

                              v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                              ln_x2 = quantile("Normal",v,6.86,1.46); /* trunk. NF från (1.4.4) på [min,max] */
                              xk_2=exp(ln_x2);
                              yk_2=0;ln_y2=0;
                      end;
                      /*1.4.3*/
                      if I_y2=1 then do;
                            /* Ungtall */
                             a=1*5.21;/*(1.4.5) för min=5.21 */
                             Fa=cdf("Normal",a,8.77,1.36);
                             b=1*10.78;/*(1.4.5) för max=10.78 */
                             Fb=cdf("Normal",b,8.77,1.36);

                             v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                             ln_x2 = quantile("Normal",v,8.77,1.36); /* trunk. NF från (1.4.5) på [min,max] */
                             xk_2=exp(ln_x2);

                /* Skaddade ungtallar */
*               e_k=rand("Normal",0,0.95);* regression (1.4.5);
*               ln_y2 = 0.7782*ln_x2 + e_k;
                             hat_y2=0.7782*ln_x2;
                             a=1*4.73;/*(1.4.5) för min=4.73*/
                             Fa=cdf("Normal",a,hat_y2,0.95);
                             b=1*8.89;/*(1.4.5) för max=8.89*/
                             Fb=cdf("Normal",b,hat_y2,0.95);

                             v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                             ln_y2 = quantile("Normal",v,hat_y2,0.95); /* trunk. NF från (1.4.5) på [min,max] */

                             if ln_y2>ln_x2 then do;
                              y2_ok=0;
                              ln_y2=(0.8 + 0.2*rand("Uniform"))*ln_x2;*kan behövas en justering F(a)=0.8 F(b)=1;
                             end;
                             yk_2=exp(ln_y2);
                      end;/*1.4.3*/
             end;/*I_x2=1*/
           end;/*I_x1=0*/

           if I_x1=1 then do;
                     /*1.4.6-9*/
                     /* parametrar är baserade på SIM-populationen: Pop_1535 */
             if I_x2=1 then do;
                              /*1.4.6*/
                      if I_y1=0 and I_y2=0 then do;
                               /* Ungtall */
*               e_k=rand("Normal",0,1.152);* regression Pop_1535;
*               ln_x2 = 0.9869*ln_x1 + e_k;
                             hat_x2=0.9869*ln_x1;
                             a=1*4.71;/*(1.4.6) för min=4.71*/
                             Fa=cdf("Normal",a,hat_x2,1.152);
                             b=1*11.51;/*(1.4.6) för max=11.51*/
                             Fb=cdf("Normal",b,hat_x2,1.152);

                             v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                              ln_x2 = quantile("Normal",v,hat_x2,1.152); /* trunk. NF från ln_x1 (1.4.6) på
                                                             [min,max] */
                             xk_2 =exp(ln_x2);
                             yk_2=0;ln_y2=0;
                      end;
                              /*1.4.7*/
```

```
		if I_y1=0 and I_y2=1 then do;
				/* Ungtall */
*		e_k=rand("Normal",0,1.30);  * regression (1.4.7) i;
*		ln_x2 = 1.0653*ln_x1 + e_k; * Pop_1535;
			hat_x2=1.0653*ln_x1;
			a=1*4.37;/*(1.4.7) för ln_x1 min=4.37*/
			Fa=cdf("Normal",a,hat_x2,1.30);
			b=1*11.51;/*(1.4.7) för ln_x1 max=11.51*/
			Fb=cdf("Normal",b,hat_x2,1.30);

			v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
			ln_x2 = quantile("Normal",v,hat_x2,1.30); /* trunk. NF för ln_x1 från (1.4.7) på
							[min,max] */
			xk_2 =exp(ln_x2);

			/* Skaddade ungtallar */
*		e_k=rand("Normal",0,1.03);   * regression i (1.4.8);
*		ln_y2 = 0.7622*ln_x2 + e_k;  * för ln_y1 = 0.7622*ln_x1 + e_k;
			hat_y2=0.7622*ln_x2;
			a=1*4.47;/*(1.4.8) för ln_y1 min=4.47*/
			Fa=cdf("Normal",a,hat_y2,1.03);
			b=1*9.55;/*(1.4.8) för ln_y1 max=9.55*/
			Fb=cdf("Normal",b,hat_y2,1.03);

			v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
			ln_y2 = quantile("Normal",v,hat_y2,1.03); /* trunk. NF från (1.4.8) på [min,max] */

			if ln_y2>ln_x2 then do;
			  y2_ok=0;
			  ln_y2=(0.8 + 0.2*rand("Uniform"))*ln_x2;*kan behövas en justering F(a)=0.8 F(b)=1;
			end;
			yk_2=exp(ln_y2);
		end;/*1.4.7*/
			/*1.4.8*/
		if I_y1=1 and I_y2=0 then do;
				/* Ungtall */
*		e_k=rand("Normal",0,1.27);  * regression (1.4.8) i;
*		ln_x2 = 0.9082*ln_x1 + e_k; * Pop_1535;
			hat_x2=0.9082*ln_x1;
			a=1*4.87;/*(1.4.8) för ln_x1 min=4.87*/
			Fa=cdf("Normal",a,hat_x2,1.27);
			b=1*11.6;/*(1.4.8) för ln_x1 max=11.6*/
			Fb=cdf("Normal",b,hat_x2,1.27);

			v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
			ln_x2 = quantile("Normal",v,hat_x2,1.27); /* trunk. NF för ln_x1 från (1.4.8) på
							[min,max] */
			xk_2 =exp(ln_x2);
			yk_2=0;ln_y2=0;

		end;
	/*1.4.9*/
		if I_y1=1 and I_y2=1 then do;
				/* Ungtall */
*		e_k=rand("Normal",0,0.89);  * regression (1.4.9) i;
*		ln_x2 = 1.0068*ln_x1 + e_k; * Pop_1535;
			hat_x2=1.0068*ln_x1;
			a=1*4.65;/*(1.4.9) för ln_x1 min=4.65*/
			Fa=cdf("Normal",a,hat_x2,0.89);
			b=1*11.28;/*(1.4.9) för ln_x1 max=11.28*/
			Fb=cdf("Normal",b,hat_x2,0.89);

			v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
			ln_x2 = quantile("Normal",v,hat_x2,0.89); /* trunk. NF för ln_x1 från (1.4.9) på
							[min,max] */
```

```
                        xk_2 =exp(ln_x2);

            /* Skaddade ungtallar */
*           e_k=rand("Normal",0,1.25);   * regression (1.4.9) i;
*           ln_y2 = 0.9738*ln_y1 + e_k;  * Pop_1535;
                        hat_y2=0.9738*ln_y1;
                        a=1*2.7;/*(1.4.9) för ln_y1 min=2.7*/
                        Fa=cdf("Normal",a,hat_y2,1.25);
                        b=1*9.69;/*(1.4.9) för ln_y1 max=9.69*/
                        Fb=cdf("Normal",b,hat_y2,1.25);

                        v= Fa + (Fb-Fa)*rand("Uniform"); /* V~U[F(a),F(b)] */
                        ln_y2 = quantile("Normal",v,hat_y2,1.25); /* trunk. NF från (1.4.9) på [min,max] */

                         if ln_y2>ln_x2 then do;
                          y2_ok=0;
                          ln_y2=(0.8 + 0.2*rand("Uniform"))*ln_x2;*kan behövas en justering F(a)=0.8 F(b)=1;
                         end;
                         yk_2=exp(ln_y2);
                  end;/*1.4.9*/

                 end;/*I_x2=1*/

                end;/*I_x1=1*/

   end;/*1.4*/

run;
```