



Norges miljø- og  
biovitenskapelige  
universitet

Masteroppgave 2017 60 stp  
Institutt for Husdyr- og Akvakulturvitenskap

## **Metoder for beregning av innavl og effektiv populasjonsstørrelse basert på slektskapsinformasjon fra Norske varmbloodstravere**

Methods for estimating inbreeding and effective population size using pedigree data from Norwegian standardbred trotters

Christine Seljenes Storhaug  
Master i husdyrvitenskap

## Forord

Denne masteroppgaven ble skrevet ved Institutt for Husdyr- og Akvakulturvitenskap (IHA) ved Norges miljø- og biovitenskaplige universitet (NMBU). Dette er en avsluttende oppgave ved masterstudiet i husdyrvitenskap med fordypning i avl og genetikk.

Interessen for avl og genetikk har kommet gradvis gjennom bachelorgraden, og da var det ikke tvil om hvilken retning som skulle velges da jeg bestemte meg for å ta en mastergrad. Når tiden var inne for å velge oppgave ville jeg skrive på hest, og da helst varmblodstraver. Da denne oppgaven ble forslått ble den fort min, å kunne skrive på et tema som man interesserer seg for er helt optimum. Tiden fra oppgaven ble valgt til nå har vært lang, med både oppturer og nedturer står jeg her med oppgaven i hånden.

En stor takk til mine veiledere, Gunnar Klemetsdal og Hanne F. Olsen for god veiledning, gode innspill, god læring og konstruktiv kritikk gjennom hele oppgaven. En ekstra takk til Hanne F. Olsen for hjelpen med programmet SAS når jeg trengte det som mest. Jeg vil også takke Det Norske Travselskapet for tilgangen til datasettet.

Vil få takke familien min som alltid har hatt troen på meg, og som er grunnen til at jeg står her i dag. Mamma og Pappa, dere er best. Jeg vil også takke min kjære for å ha holdt ut med meg i tiden hvor oppgaven slet som mest på meg, takk for at du er du og for at du har støttet meg gjennom det hele.

Jeg vil også rette en stor takk til venner, som har stilt opp og gitt meg moralsk støtte gjennom studiet. I tillegg vil jeg takke Stine Samsonstuen for korrekturlesning, og får gode råd på veien til målet.

Institutt for husdyr- og akvakulturvitenskap

NMBU

Ås, 13. 03.2017

.....

Christine S. Storhaug

## Sammendrag

Effektiv populasjonsstørrelse ( $N_e$ ) kan estimeres på forskjellige måter, de vanligste metodene for estimering av  $N_e$  er å benytte slektskapsinformasjon. Formålet med denne oppgaven er å teste ut fem forskjellige metoder for å estimere  $N_e$  ved bruk av slektskapsinformasjon fra norsk-registrerte varmblodstravere. De fem metodene baserte seg på individuelle innavlskoeffisienter til dyrene i referansepopulasjonen og deres foreldre (Metode 1), logregresjon hvor  $\ln(1-F)$  ble regresset på fødselsår (Metode 2), på individuell innavlskoeffisient og komplett generasjonsekvivalent (CGE) (Metode 3), logregresjon hvor  $\ln(1-F)$  ble regresset på CGE (Metode 4) og enkel lineær regresjon hvor individuell  $F$  ble regresset på fødselsår (Metode 5).

Den originale slektskapsfilen bestod av 90559 varmbloidshester fått fra Det Norske Travelskapet. Datasettet i denne oppgaven består av en slektskaps-fil genererte fra 5000 hester født fra 2010 til og med 2015 som ble definert som referansepopulasjon. Det ble hentet informasjon i maksimalt 23 generasjoner og den genererte slektskapsfilen ble bestående av 19056 hester. Ved metoden basert på enkel lineær regresjon ble en revidert utgave av den originale slektskapsfilen ( $N=90481$ ) brukt til estimering av  $N_e$ .

Gjennomsnittlig slektskapsdybde i referansepopulasjonen var på 9,86 generasjoner, og generasjonsintervallene ble beregnet til 11,62 og 11,72. Innavlskoeffisienter ( $F$ ) ble estimert for individene i den genererte slektskapsfilen som legger grunnlaget for beregningene av innavlsraten i metodene. Resultatene viser stor variasjon ved bruk av de forskjellige metodene for beregning av  $N_e$ .  $N_e$  varierer fra 21 til 128 individer, hvor Metode 4,  $\ln(1-F)$  regresset på CGE gir minst  $N_e$  og Metode 1 som baseres på innavlskoeffisientene til referansepopulasjonen og deres foreldre gir størst  $N_e$ .

Bruk av CGE-verdier i populasjonen for varmbloidsstravere kan forårsake over og underestimering av  $\Delta F$  ved bruk av Metode 3 og 4. Det bør unngås å bruke generasjonsintervall i beregninger av  $N_e$ , fordi lange generasjonsintervall (11,60 år) vil påvirke  $\Delta F$  som kan føre til underestimering av  $N_e$ . Logregresjon på fødselsår mot bruk av lineær regresjon på fødselsår viser at logregresjon gir bedre resultater grunnet feilmarginsverdiene i datasettet er betydelig redusert. Metode 1 og 2 vil sammen kunne gi en god vurdering av genetisk variasjon i populasjonen ut ifra en praktisk vinkling.

## Abstract

There are different ways to estimate effective population size ( $N_e$ ), where pedigree is the usual method for this estimation. The purpose of this task is to test five different methods of estimating  $N_e$  with the use of pedigree data from Norwegian-registered Standardbred trotters. The five methods are based on individual inbreeding coefficients for the animals in the reference population and their parents (Method 1), log-regression where  $\ln(1-F)$  was regressed on the year of birth (Method 2), individual inbreeding coefficients and complete generation equivalent (CGE) (Method 3), log-regression where  $\ln(1-F)$  was regressed on CGE (Method 4) and simple linear regression where individual  $F$  were regressed on the year of birth (Method 5).

The original pedigree from “Det Norske Travelskapet” consisted of 90559 standardbred trotters. The dataset in this task consist of a pedigree-file generated from 5000 horses born from 2010 to and including 2015 which were defined as the reference population. Information were gathered from maximum 23 generations and the generated pedigree-file now consisted of 19056 horses. The method based on simple linear regression used a revised edition of the original pedigree ( $N=90481$ ) used to estimate  $N_e$ .

Mean pedigree depth in the reference population was 9,86 generations, and the generation intervals were estimated to be 11,62 and 11,72. The basis for estimating the rate of inbreeding used in the different methods comes from individual inbreeding coefficients in the generated pedigree-file. The results show a large variation on the use of the methods to estimate  $N_e$ .  $N_e$  varies from 21 to 128 individuals where Method 4 ( $\ln(1-F)$ ) regressed on CGE have the lowest estimate of  $N_e$ , and Method 1 based on inbreeding coefficients of the reference population and their parents show the highest value of  $N_e$ .

The use of CGE-values in the Standardbred trotters population could cause over- and underestimation of  $\Delta F$  with the use of Method 3 and 4. In estimating  $N_e$  long generation intervals (11,60 years) could influence  $\Delta F$  which could lead to underestimation of  $N_e$ , and it should be avoided. Log-regression on birthyears against the use of linear regression on birthyears show that the log-regression gives better results due to the margins of error in the dataset to be substantially reduced. Method 1 and 2 together should give a good estimate of genetic variation in a population from a practical perspective.

# Innholdsfortegnelse

Forord.....	1
Sammendrag.....	2
Abstract .....	3
Innholdsfortegnelse .....	4
1. Innledning .....	5
2. Varmblodstraver i Norge.....	6
2.1 Innavl .....	6
3. Effektiv populasjonsstørrelse .....	7
4. Material og metode .....	9
4.1 Beskrivelse og bearbeiding av rådata.....	9
4.3 Beregning av raseandel franske aner i varmbloedstraveren på individ nivå.....	10
4.4 Beregning av innavlskoeffisienter basert på slektskap .....	10
4.5 Komplette generasjonsekvivalent (CGE).....	11
4.6 Generasjonsintervall.....	11
4.7 Effektiv populasjonsstørrelse - Metode 1 .....	12
4.8 Effektiv populasjonsstørrelse - Metode 2 .....	13
4.9 Effektiv populasjonsstørrelse - Metode 3 .....	14
4.10 Effektiv populasjonsstørrelse - Metode 4 .....	15
4.11 Effektiv populasjonsstørrelse - Metode 5 .....	16
5. Resultat.....	17
5.1 Raseandeler .....	18
5.3 Generasjonsintervall.....	20
5.4 Regresjonsanalyse.....	21
5.5 Beregning av effektiv populasjonsstørrelse .....	23
6. Diskusjon .....	24
6.1 Raseandeler .....	24
6.2 Genetisk fremgang og generasjonsintervall .....	25
6.3 Slektskapsdybde i datasettet.....	26
6.4 Forskjeller i metoder for beregning av $N_e$ .....	27
7. Konklusjon .....	30
Referanser .....	31

## 1. Innledning

Antall hester har variert siden 1900-tallet da hesten var det store arbeidsjernet i landbruksområder og transportmiddelet rundt om i hele landet. Hesteantallet ble drastisk redusert i årene 1943-1980 da mekaniseringen av landbruket startet for fullt (Vangen 2016a). Hesten har gått fra å være et hjelpemiddel, som landbrukshest og transportmiddel til å bli et sports- og familiedyr, og har etter mekaniseringen fått flere nye bruksområder som blant annet trav, galopp, ridning og generelt fritidsbruk (Econ 2009). En forskningsrapport viste at antall hester sakte har bygd seg oppover og ble indikert til 125 000 i 2012 av Vik og Farstad (2012) ved deres rapport på Hest, hestehold og føring. Rapporten viste utbredelsen av raser i Norge hvor varmbloodstraveren utgjorde 25 %.

Travhesten består av to forskjellige raser, kaldbloodstraver og varmbloodstraver. Kaldbloodstraveren er en av de nasjonale hesterasene til Norge, mens varmbloodstraveren ikke er av norsk opprinnelse. Det Norske Travelskapet (DNT) ble stiftet i 1875, og har siden den gang hatt ansvaret sammen med Norsk Hestesenter for de to travrasene i Norge. Tidligere ble både kaldblods- og varmbloodstraveren skrevet inn i felles stambok, men i 1903 vedtok DNT en lovendring hvor varmbloodige og kaldbloodige ikke lenger skulle ha felles stambok. Stamboken ble delt i to deler, hvor det i del A var hester av rene norske aner og del B var de varmbloodige hestene både rene og krysninger stambokført (Norsk Hestesenter 2003). Norsk hestesenter har sammen med DNT siden 1995 utarbeidet avlsplan for begge travrasene, i Januar 2015 overtok DNT det fulle ansvaret for utarbeidelse av avlsplanene (Det Norske Travelskapet 2015). Avlsplanene inneholder retningslinjer for driften av avlsarbeidet, og hvordan avlsarbeidet bør drives i fremtiden.

Det er tidligere gjort arbeid på innavl og effektiv populasjonsstørrelse på kaldbloodstraveren i Norge (Olsen & Klemetsdal 2009), derimot er det gjort lite på den norske varmbloodstraveren. Det er ikke tidligere blitt gjort arbeid på hele populasjonen av varmbloodstraver, men i arbeidet til Samsonstuen (2015) hvor det ble sett på effekt av innavl på prestasjonsegenskaper basert på to årganger født 1988 og 2006/2007, ble det også beregnet effektiv populasjonsstørrelse.

Det finnes mange metoder for å undersøke innavl og effektiv populasjonsstørrelse ved bruk av slektskapsinformasjon. Det ville derfor vært interessant å teste forskjellige metoder som ofte blir brukt til estimering av innavl og effektiv populasjonsstørrelse ved bruk av den norske besetningen av varmbloodstravere.

## 2. Varmblodstraver i Norge

Den norske populasjonen av varmblodstraver blir karakterisert som en krysningsrase og består i hovedsak av amerikansk standardbred og fransk traver, med små innslag av den russiske Orlowertraveren (Norsk Hestesenter 2003). På 1900-tallet ble de første hestene importert, den første varmblodshingsten Jordan ble importert fra USA og den russiske Orlowerhoppa Secunda ble importert fra Tyskland. Disse hestene og deres avkom danner grunnlaget for den Norske rasebestanden av varmblodstravere (Norsk Hestesenter 2003).

Formålet med avlsarbeidet på varmblodstraver er å videreutvikle internasjonale konkurransedyktige travhester og oppnå genetisk fremgang for travprestasjon, holdbarhet, eksteriør, håndterbarhet/lynne, bevegelser i trav og fruktbarhet (Det Norske Travselskapet 2015). Etter at tvungen kåring av avlshingster ble opphevet i 2015 (Vangen 2016b), er det kun avlsvurdering av hingster som brukes i avl. Avlsvurderingen er en informasjonskilde for oppdrettere, slik at de skal få mest mulig informasjon om hingsters egenskaper før valg av hingst. En hingst kan brukes i avl så lenge den ikke har diskvalifiserende defekter som pungbrokk, urhingst, klapphingst, bittfeil og alvorlige testikkelfeil (Det Norske Travselskapet 2015).

### 2.1 Innavl

Innavl på varmblodstraver har ikke tidligere vært sett på som et stort problem, på grunn av importeringen av amerikanske- og franske travere. Dette gjør at valgmulighetene ved seleksjon av dyr til avl er stor ettersom man kan se til andre land for å få materialet man ønsker, kravet DNT har satt i forhold til å minske innavl er en bedekningskvote på 150 hopper per hingst (Norsk Hestesenter 2003).

Innavl skjer når to individer som er i slektskap med hverandre gjennom felles forfedre pares (Falconer & Mackay 1996). Graden av innavl avhenger av hvor nært slektskapet er og størrelsen på populasjonen og blir beregnet ved hjelp innavlskoeffisienter (F). Innavlskoeffisientene til et individ måler hvor stor sannsynligheten er for at de to allelene som arves fra far og mor er identiske i opphav (Falconer & Mackay 1996).

I arbeidet til Moureaux et al. (1996) på fransk traver fant de gjennomsnittlig innavlskoeffisient for avkom født 1989 til 1992 ( $n=45\ 754$ ) på 1, 86% og kun 1% av avkommene hadde ingen felles forfedre derav ikke innavlet. MacCluer et al. (1983) viste i sitt arbeid at de amerikanske travhestene som hadde komplett slektskap var innavlet i de første fem generasjonene og avkommene født 1980-1983 hadde en gjennomsnittlig

innavlskoeffisient estimert til rundt 8%. Oppgaven til Samsonstuen (2015) viste at den gjennomsnittlig innavlskoeffisienten til norsk-registrerte varmblodstravere født 1988 og 2006/2007 var på 6,1%.

### **3. Effektiv populasjonsstørrelse**

Den effektive populasjonsstørrelsen ( $N_e$ ) blir definert som det antall individer som gir grunnlag for den beregnede innavlsraten ( $\Delta F$ ) eller endringen av variasjon i genfrekvenser, dersom det avles i en ideell populasjon (Falconer & Mackay 1996). Den effektive populasjonsstørrelsen er anbefalt av FAO (1998) til å være 50 dyr eller mer, på bakgrunn av den genetiske variasjonen i en populasjon. Dersom  $N_e$  er mindre bør det settes i gang tiltak for å øke den effektive populasjonsstørrelsen for at populasjonen ikke skal utryddes og miste genetisk.  $N_e$  er direkte koblet til innavlsrate og variasjonen som mistes ved genetisk drift i en populasjon (Falconer & Mackay 1996), og kan derfor brukes som en indikator i bevaringsarbeid for utrydningstruede populasjoner.

Det er ingen standard metode for beregningen av  $N_e$ , men de vanligste metodene i husdyravl legger grunnlaget på slektskapsinformasjon i populasjonen. En metode introdusert av Wright (1931) baserer seg på kjønnsfordelingen i populasjonen som undersøkes. Dette er en enkel metode og det trengs kun informasjon om hvor mange avlshingster og avlshopper det er i referansepopulasjonen.

Cervantes et al. (2011) introduserte en metode for beregning av  $N_e$  ved bruk av «coancestry» som baserer seg på forholdet mellom to foreldre og er identisk med innavlskoeffisienten til deres avkom dersom de pares (Falconer & Mackay 1996). Metoden er modifisert fra Gutierrez et al. (2009) og benytter «coancestry» i stedet for individuelle innavlskoeffisienter sammen med den komplette generasjonsekvivalenten.

Ved å bruke slektskapsinformasjon kan man estimere innavlskoeffisienter ( $F$ ) og innavlsraten ( $\Delta F$ ) i populasjonen og deretter beregne  $N_e$ .

En metode for beregning av  $N_e$  ble introdusert i arbeidet til Olsen et al. (2010) hvor de så på slektskapsstruktur og genetisk variasjon i Dølahesten og Nordland/Lyngshesten. Denne metoden benytter en referansepopulasjon og beregner den effektive populasjonsstørrelsen ved bruk av individet i referansepopulasjonens  $F$  og foreldrenes  $F$ . Metoden gir en indikasjon på hvordan populasjonen ser ut per dagsdato.



Flere av metodene i denne oppgaven benytter regresjon til estimering av  $N_e$ . Enkel lineær regresjon er en alminnelig metode for å estimere  $\Delta F$ . Ved enkel lineær regresjon er målet å finne hvilken lineær sammenheng det er mellom  $Y$ -variabelen og  $X$ -variabelen (Løvås 2010). Standardmodellen for regresjon er:

$$Y = \beta_0 + \beta_1 X + e_i$$

Hvor  $Y$  er responsen (avhengig variabel),  $X$  er uavhengig variabel,  $\beta_0$  er konstantleddet,  $\beta_1$  er regresjonsvariabelen til  $X$  og  $e_i$  er feilleddene i modellen.

I enkel lineær regresjon antas det at verdien på  $Y$  ikke kan være høyere enn 1 eller lavere enn 0, som kan føre til flere feilledd ved bruk av  $Y$ -verdier som ikke er enten 1 eller 0. Ved å bruke logregresjon og den naturlige logaritmen til et tall, kan man omforme koeffisientene i ikke-lineære variabler slik at de får en lineær sammenheng og kan brukes i regresjonsanalysen (Løvås 2010). Dette gjør at logregresjonen har fått en stor rolle i estimeringen av  $N_e$ .

## 4. Material og metode

### 4.1 Beskrivelse og bearbeiding av rådata

Slektskapsfilen som ble brukt i denne oppgaven kommer fra DNT. Filen bestod av 90559 individer med fødselsår fra 1822 til og med 2015. Individene var registrert med fullt navn, identitetsnummer og fødselsdato. I tillegg inneholdt filen foreldreinformasjon på individene samt fødselsdato på foreldrene. Det ble det gjort mye arbeid med å rette opp den originale slektskapsfilen fra DNT for feil og mangler. Gjennom dette arbeidet ble det gjort endringer som skapte komplikasjoner for det videre arbeidet, og det ble bestemt å bruke en tidligere bearbeidet fil fra samme originalfil gjort av Hanne F. Olsen.

Filen ble bearbeidet ved hjelp av Excel (Microsoft Excel 2016). Slektskapsinformasjonen til individene med ukjent avstamming ble satt til 0 og ble definert som foundere. Hel-founder er definert som individer med ukjente foreldre, og individer hvor kun en av foreldrene var kjent ble definert som halv-foundere (Boichard 1997). Tre hester ble slettet på grunn av feilregistrering eller dobbeltregistrering. I tillegg inneholdt slektskapsfilen fiktive foreldre med registreringsnummer "999997" og "999998", disse individene ble satt til null. Etter bearbeiding inneholdt slektskapsfilen 90556 individer.

Referansepopulasjonen ble definert som de seks siste årgangene fra og med 2010 til og med 2015. Fem hester ble slettet på grunn av ukjent avstamming og to hester fikk rettet opp feil fødselsår. Antall hester i referansepopulasjonen etter bearbeiding ble 5000 (se Tabell 1).

Det ble nøstet en slektskaps-fil fra referansepopulasjonen på 19056 dyr ved hjelp av programmet SAS 9.4 (SAS 2013). Det ble sporet slektskap så langt tilbake det forelå informasjon i generasjonene. Dyr som manglet fødselsår fikk deretter tildelt fiktive fødselsår hvor kriteriet var at foreldrene skulle være minst ett år eldre enn sine avkom. ID-numrene i slektskapsfilen ble renummerert, individene ble sortert etter årstall og duplikater ble fjernet. I tillegg ble det hentet ut informasjon om antall foundere og slektskapsdybden til referansepopulasjonen ble beregnet.

Tabell 1. Fordeling av hopper, hingster og vallaker i referansepopulasjonen.

	Hingster	Hopper	Vallaker	Total
2010	286	633	450	1369
2011	238	352	510	1100
2012	241	406	180	827
2013	242	325	52	619
2014	293	287	13	593
2015	243	249	0	492
Total	1543	2252	1205	5000

#### **4.3 Beregning av raseandel franske aner i varmblodstraveren på individ nivå.**

Programmet SAS 9.4 ble brukt til beregning av de franske raseandelene i referansepopulasjonen, og prosedyren som ble brukt stammer fra et tidligere arbeid gjort av Lykkjen et al. (2012).

Individer med franske aner ble sortert ut og fransk raseandel ble beregnet ut fra slektskapsfilen. Det ble satt en restriksjon ved sorteringen av individene, og det ble kun sortert ut individer hvor ID-nummeret startet på «F». Deretter ble individer som hadde ID-nummer 'FI', 'FF' og 'FO' fjernet da disse ID-numrene ble vurdert til å ikke være franske individer.

#### **4.4 Beregning av innavlskoeffisienter basert på slektskap**

Graden av innavl hos et individ måles ved hjelp av innavlskoeffisienten (F). F beskriver sannsynligheten for at de to allelene som arves fra far og mor er identiske i opphav (Falconer & Mackay 1996). Den genererte slektskapsfilen ble brukt til å estimere individuelle innavlskoeffisienter (F) på individene i slektskapsfilen ved hjelp av SAS 9.4. Innavlskoeffisientene ble beregnet ved bruk av modellen:

$$F_x = \sum \left(\frac{1}{2}\right)^n (1 + F_A) \quad (\text{Falconer \& Mackay 1996})$$

hvor  $F_x$  er innavlskoeffisienten,  $n$  er antall individer og  $F_A$  er innavlskoeffisienten til individet A.

#### 4.5 *Komplett generasjonsekvivalent (CGE)*

Den komplette generasjonsekvivalenten baserer seg på andel kjente forfedre per generasjon. Det dannes en gjennomsnittlig verdi for individene i referansepopulasjon basert på antall generasjoner hvor forfedrene er kjent.

Det ble beregnet individuell CGE for referansepopulasjonen ved hjelp av SAS 9.4 basert på formelen for CGE:

$$CGE_i = \sum_{j=1}^{n_j} \frac{1}{2^{g_{ij}}} \quad (\text{Boichard 1997})$$

hvor  $n_j$  er det totale antall forfedre til dyr  $j$ , og  $g_{ij}$  er antall generasjoner mellom dyr  $j$  og dens forfeder  $i$ .

#### 4.6 *Generasjonsintervall*

Generasjonsintervallet ( $L$ ) beregnes som gjennomsnittsalderen på foreldrene når selekterte avkom blir født (Falconer & Mackay 1996).

Generasjonsintervallet ble beregnet ved hjelp av SAS 9.4. Det ble satt en restriksjon på foreldrene som skulle inkluderes i generasjonsintervallet var minimum 4 år og maksimum 25 år. Minimumsrestriksjonen ble satt på grunn av de fiktive fødselsårene og maksimumrestriksjonen ble satt på grunn av opplagte feil i fødselsår hvor foreldrene var flere tiår eldre enn avkommene. Generasjonsintervallet ble beregnet separat for mor og far, og i tillegg ble det beregnet et gjennomsnittlig generasjonsintervall for foreldrene.

Generasjonsintervallet ble beregnet ved hjelp av modellene:

$$L_{far} = \frac{1}{n} \sum_{i=1}^n Faralder_i$$

$$L_{mor} = \frac{1}{n} \sum_{i=1}^n Moralder_i$$

$$\bar{L} = \frac{L_{Far} + L_{Mor}}{2}$$

Hvor  $L_{far}$  er generasjonsintervallet til fedrene,  $n$  er antall fedre og  $faralder_i$  er alder på fedre.  $L_{mor}$  er generasjonsintervallet til mødrene,  $n$  er antall mødre og  $moralder_i$  er alder på mødre.  $\bar{L}$  er gjennomsnittlig generasjonsintervall basert på generasjonsintervallet til far og mor.

#### 4.7 Effektiv populasjonsstørrelse - Metode 1

Metode 1 er basert på Olsen et al. (2010), og baseres på individuelle innavlskoeffisienter til dyr i referansepopulasjonen (n=5000). Den gjennomsnittlige innavlsraten til referansepopulasjonen gir grunnlaget for beregning av effektiv populasjonsstørrelse ( $N_e$ ) ved bruk av de seks siste årgangene i populasjonen.

Den effektive populasjonsstørrelsen ble beregnet ved hjelp av modellen:

$$N_e = \frac{1}{2\overline{\Delta F}_t}$$

hvor  $\overline{\Delta F}_t$  ble beregnet ved bruk modellen (Olsen et al. 2010):

$$\overline{\Delta F}_t = \frac{1}{N} \sum_{i=1}^N \frac{F_i - \frac{1}{2}(F_{sire_i} + F_{dam_i})}{1 - \frac{1}{2}(F_{sire_i} + F_{dam_i})}$$

Hvor N er antall individer,  $F_i$  er den individuelle innavlskoeffisienten for dyr  $i$  i referansepopulasjonen,  $F_{sire_i}$  og  $F_{dam_i}$  er de individuelle innavlskoeffisientene til foreldrene til dyr  $i$  i referansepopulasjonen.

Ved beregning av standardfeil for  $N_e$  til Metode 1 ble modellen til Cervantes et al. (2008) benyttet som er modifisert fra Gutiérrez et al. (2008):

$$SE(\overline{N}_e) = 2\overline{N}_e^2 \sigma_{\Delta F} \frac{1}{\sqrt{\overline{N}_e}}$$

hvor  $SE(\overline{N}_e)$  er standardfeilen beregnet for den effektive populasjonsstørrelsen ( $N_e$ ) og  $\sigma_{\Delta F}$  er standardavviket for innavlsraten ( $\Delta F$ ).

#### 4.8 Effektiv populasjonsstørrelse - Metode 2

Metode 2 er basert på arbeidet til Pérez-Enciso (1995), og beregnes ved bruk av den nøytede slektskapsfilen (N=19056). Denne metoden for å beregne effektiv populasjonsstørrelse er basert på logregresjon av  $\ln(1-F)$  ved bruk av modellen:

$$\ln(1 - F_i) = y = \beta_0 + \beta_1 * f_{odt} + e_i$$

hvor  $\ln(1-F_i)$  er den naturlige logaritmen til (1- individuell innavlskoeffisient (F) til dyr  $i$ ),  $\beta_0$  er konstant og  $\beta_1$  er regresjonsvariabelen av fødselsår (fodt) på  $y$ .

$\Delta F$  og  $N_e$  ble beregnet ved hjelp av modellene:

$$\Delta F = 1 - e^{\beta_1} \text{ og } N_e = \frac{1}{2\Delta F \times L}$$

hvor  $L$  er det gjennomsnittlige generasjonsintervallet til foreldrene.

Minimums- og maksimumsverdi for  $N_e$  for Metode 2 ble beregnet ved hjelp av modellene:

$$\Delta F(\max) = 1 - e^{\beta_1 - SE_{\beta_1}}$$

$$\Delta F(\min) = 1 - e^{\beta_1 + SE_{\beta_1}}$$

$$N_{e\min} = \frac{1}{2\Delta F(\max) \times L} \text{ og } N_{e\max} = \frac{1}{2\Delta F(\min) \times L}$$

hvor  $SE$  er standardfeilen til  $\beta_1$  og  $L$  er generasjonsintervallet.

#### 4.9 Effektiv populasjonsstørrelse - Metode 3

Metode 3 er basert på Gutierrez et al. (2009) som er modifisert fra Gutierrez et al. (2008), og blir beregnet ved bruk av referansepopulasjonen (N=5000). Metoden baseres på individuell innavlskoeffisient og individuell komplett generasjonsekivalent. Beregningene ble gjort i programmet SAS 9.4.

Modellen som ble brukt til beregning av innavlsraten ( $\overline{\Delta F}$ ):

$$\overline{\Delta F}_i = 1 - {}^{CGE_i-1}\sqrt{(1 - F_i)} \quad (\text{Gutierrez et al. 2009})$$

hvor  $\overline{\Delta F}_i$  er gjennomsnittlig innavlsrate til dyr  $i$ ,  $CGE_i$  er den individuelle komplette generasjonsekivalenten til individ  $i$  og  $F_i$  er den individuelle innavlskoeffisienten til individ  $i$ .

Den effektive populasjonsstørrelsen ble beregnet ved hjelp av modellen:

$$\overline{N}_e = \frac{1}{2\overline{\Delta F}_i}$$

Ved beregning av standardfeil for  $N_e$  til Metode 3 ble modellen til Cervantes et al. (2008) benyttet, som er modifisert fra Gutiérrez et al. (2008):

$$SE(\overline{N}_e) = 2\overline{N}_e^2 \sigma_{\Delta F} \frac{1}{\sqrt{\overline{N}_e}}$$

hvor  $SE(\overline{N}_e)$  er standardfeilen beregnet for den effektive populasjonsstørrelsen ( $N_e$ ),  $\sigma_{\Delta F}$  er standardavviket for innavlsraten ( $\Delta F$ ).

#### 4.10 Effektiv populasjonsstørrelse - Metode 4

Metode 4 er basert på Hillestad et al. (2015) og blir beregnet ved bruk av referansepopulasjonen (N=5000) som datagrunnlag. Metoden baseres på logregresjon, hvor  $\ln(1-F_i)$  regresses på komplett generasjonsekivalent (CGE) ved bruk av modellen:

$$\ln(1 - F_i) = y = \beta_0 + \beta_1 * CGE + e_i$$

hvor  $\ln(1-F_i)$  er lik  $y$ ,  $\beta_0$  er konstant,  $\beta_1$  er regresjonsvariabelen til CGE på  $y$  og  $e_i$  er residualene (feilene).

Innavlsraten ( $\Delta F$ ) og den effektive populasjonsstørrelsen ( $N_e$ ) ble beregnet ved hjelp av følgende formler:

$$\Delta F = 1 - e^{\beta_1} \text{ og } N_e = \frac{1}{2\Delta F}$$

Minimums- og maksimumsverdi for  $N_e$  for Metode 4 ble beregnet ved hjelp av modellene:

$$\Delta F(max) = 1 - e^{\beta_1 - SE\beta_1}$$

$$\Delta F(min) = 1 - e^{\beta_1 + SE\beta_1}$$

$$N_e min = \frac{1}{2\Delta F(max)} \text{ og } N_e max = \frac{1}{2\Delta F(min)}$$

Hvor  $\Delta F(max)$  og  $\Delta F(min)$  er grenseverdiene til innavlsraten basert på standardfeilen til  $\beta_1$ .



#### **4.11 Effektiv populasjonsstørrelse - Metode 5**

I Metode 5 ble det brukt en revidert fil av den originale slektskapsfilen som inneholdt 90481 individer, hvor hester uten ID nummer var slettet. Den reviderte filen ble bearbeidet for å sikre at alle foreldredyrene var oppført som egne individer i filen. Individer uten fødselsår som ikke ytet genetisk bidrag videre ble slettet. Individer med kjente foreldre fikk manuelt tildelt fiktive fødselsår basert på foreldrenes fødselsår, hvor det antas at foreldrene er født minst ett år før sine avkom. Antall hester i pedigree-filen etter bearbeiding var 90490 individer. Det gjennomsnittlig generasjonsintervall for foreldrene i denne pedigree-filen ble beregnet til 11,76 (se Tabell 4.)

Metode 5 er basert på enkel lineær regresjon med modellen:

$$F_i = \beta_0 + \beta_1 * f_{odt_i} + e_i$$

hvor  $F_i$  er regresjonskoeffisienten på  $f_{odt_i}$  og  $e_i$  er residualer.

I enkel lineær regresjon tilsvarer resultatet for  $\beta_1$  innavlsraten ( $\Delta F$ ) og den effektive populasjonsstørrelsen ble beregnet ved hjelp av modellen:

$$N_e = \frac{1}{2\Delta F * L}$$

hvor  $L$  er det gjennomsnittlige generasjonsintervallet til foreldrene.

Minimums- og maksimumsverdi for  $N_e$  ble beregnet ved hjelp av modellen:

$$N_{e\min/\max} = \Delta F \pm SE$$

hvor  $\Delta F$  er innavlsraten og  $SE$  er standardfeilen i regresjonsanalysen.

## 5. Resultat

Tabell 2 viser beskrivende slektskapsdybde hos referansepopulasjonen.

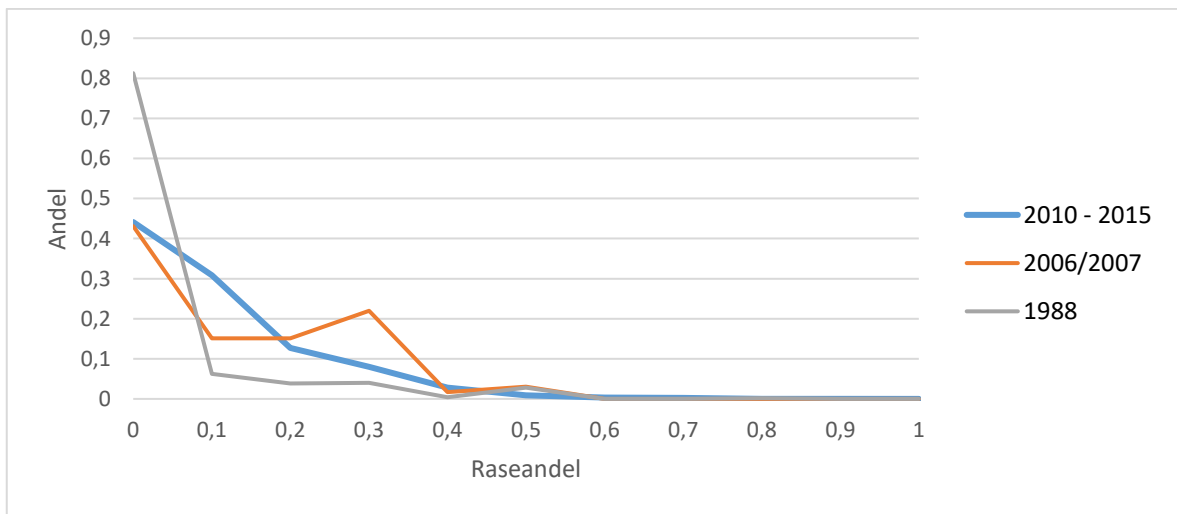
Referansepopulasjonen kan spores tilbake i maksimalt 23 generasjoner hvor CGE ligger på 9,86. Totalt antall foundere, inkludert halv-foundere er på 2460 dyr. I de fem første generasjonene er 100-95% av foreldrene kjente, mens andelen kjente foreldre synker jo lenger ut i generasjonene man kommer (Tabell 2).

Tabell 2. Beskrivende slektskapsdybde hos referansepopulasjonen (n=5000)

		Varmblodstraver
Antall individer i referansepopulasjonen		5000
Prosentandel (%) kjente foreldre i generasjon	1	100
	5	95,23
	10	61,77
	15	2,00
Totalt antall halv-foundere		1656
<i>Far missing</i>		49
<i>Mor missing</i>		1607
Totalt antall hel-foundere		804
Maksimalt antall genererte generasjoner		23
Komplett generasjonsekvivalent		9,86

## 5.1 Raseandeler

Resultatet av beregningen av individer med franske aner vises i Figur 1. Figuren er basert på tre populasjoner født 1988, 2006/2007 og referansepopulasjonen (2010-2015) i denne oppgaven. De franske raseandelene varierer mellom 19%, 57% og 56% i de tre respektive populasjonene. Populasjonene født 1988 og 2006/2007 er hentet fra arbeidet til Lykkjen et al. (2014). Figuren viser at populasjonen født i 1988 har lavere variasjon i franske andeler enn de andre populasjonene.



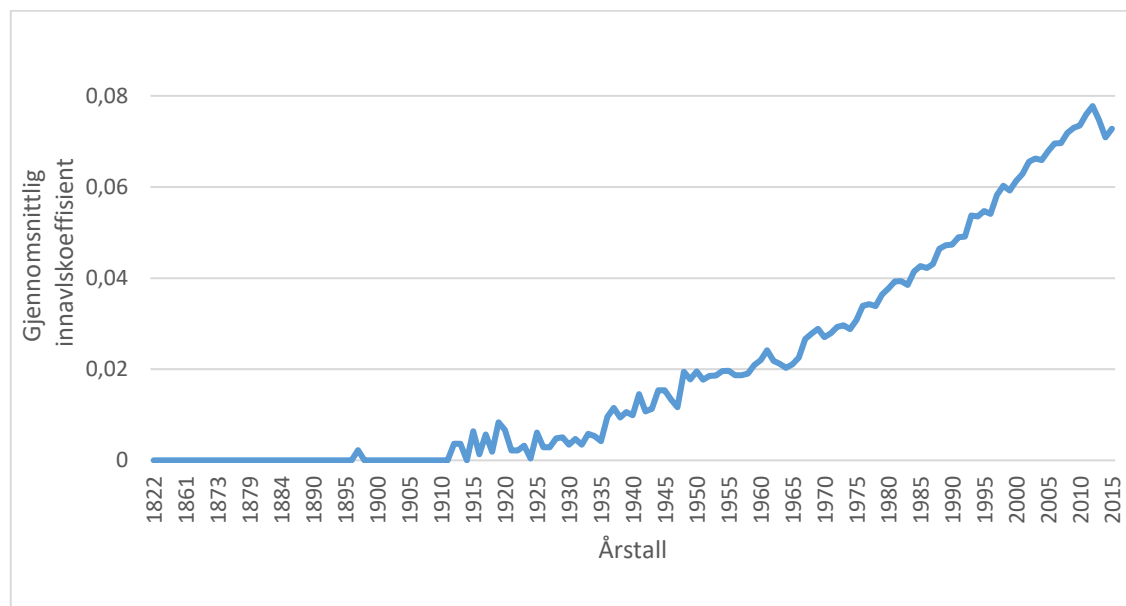
Figur 1. Viser fordelingen av franske andeler i referansepopulasjonen (født fra 2010 til og med 2015), franske andeler for individer født 2006/2007 (Lykkjen et al. 2014) og franske andeler for individer født 1988 (Lykkjen et al. 2014).

## 5.2 Gjennomsnittlige innavlskoeffisient per år

De gjennomsnittlige innavlskoeffisientene er basert på den genererte slektskapsfilen med 90490 individer som ble brukt til beregning av effektiv populasjonsstørrelse i Metode 5.

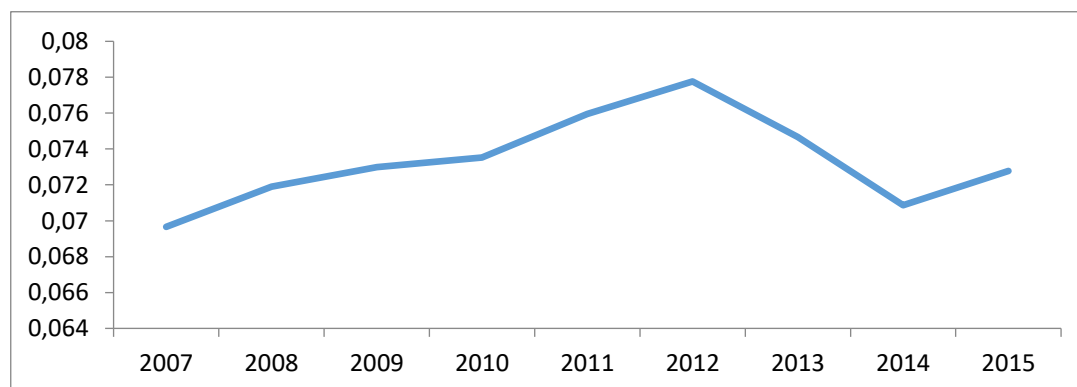
Fordelingen av gjennomsnittlige innavlskoeffisienter per år vises i Figur 2.

Innavlskoeffisientene som vises i figuren som 0 kommer av manglende slektskapsinformasjon hos individene. Innavlskoeffisientene øker gradvis fra 1900-tallet frem til 2000-tallet, og når et toppunkt i 2012 og deretter minker F de neste 4 årene.



Figur 2. Gjennomsnittlige innavlskoeffisienter per år for 90490 individer.

Figur 3 viser gjennomsnittlig innavlskoeffisient per år av årgangene fra 2007 til og med 2015. Figuren viser stigning i gjennomsnittlig innavlskoeffisienter per år frem til 2012, deretter synker innavlskoeffisientene de siste to årgangene.



Figur 3. Viser gjennomsnittlig innavlskoeffisient per år for individer født fra 2007 til og med 2015.

### 5.3 Generasjonsintervall

Tabell 3 viser resultatet fra beregningen av generasjonsintervall ved bruk av slektskapsfilen til referansepopulasjonen (N=19056). Tabellen viser separate gjennomsnitt for far og mors generasjonsintervall. Det totale gjennomsnittlige generasjonsintervallet ble brukt i beregningen for effektiv populasjonsstørrelse i Metode 2.

Tabell 3. Resultat av beregning av generasjonsintervall basert på 19056 individer.

	N	L	Std.	Min	Max
Far	17499	12,36	4,82	4,00	25,00
Mor	16257	10,85	4,18	4,00	25,00
Total	33756	11,60			

I Tabell 4 vises et gjennomsnittlig generasjonsintervall basert på den totale slektskapsfilen (N=90490). Det totale gjennomsnittlige generasjonsintervallet ble brukt til å beregne effektiv populasjonsstørrelse i Metode 5. Både Tabell 3 og 4 viser et lengre generasjonsintervall hos fedrene enn hos mødrene.

Tabell 4. Resultat av beregning av generasjonsintervall (L) basert på 90490 hester.

	N	L	Std.	Min	Max
Far	35539	12,56	4,95	4,00	25,00
Mor	28480	10,99	4,33	4,00	25,00
Total	64019	11,76			

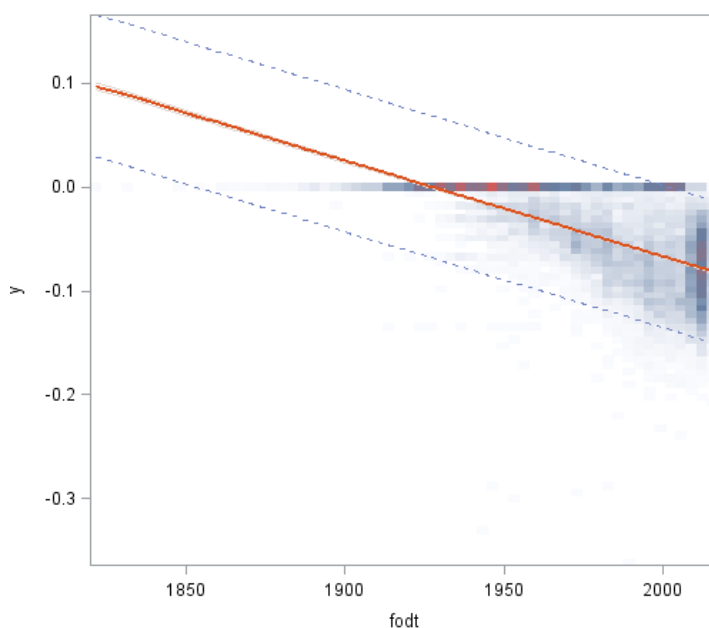
## 5.4 Regresjonsanalyse

I Tabell 5 vises resultatene fra de tre regresjonsanalysene.  $R^2$  forklarer hvor mye av variasjonen i punktene som forklares av modellene. Tabellen viser at  $R^2$  varierer noe mellom Metode 2 og 5 som er regresset på fødselsår, derimot øker  $R^2$  ved metode 4 hvor  $\ln(1-F)$  regresset på den komplette generasjonsekvivalenten (CGE). p-verdiene gir uttrykk for signifikans ved bruk av de tre regresjonsmetodene, og viser at metodene er statistisk signifikante på grunn av lave verdier.

Tabell 5. Resultat fra regresjonsanalysene for metode 2, 4 og 5. Metode 2 er  $\ln(1-F)$  regresset på fødselsår, i Metode 4 er  $\ln(1-F)$  regresset på CGE og Metode 5 er  $F$  regresset på fødselsår.

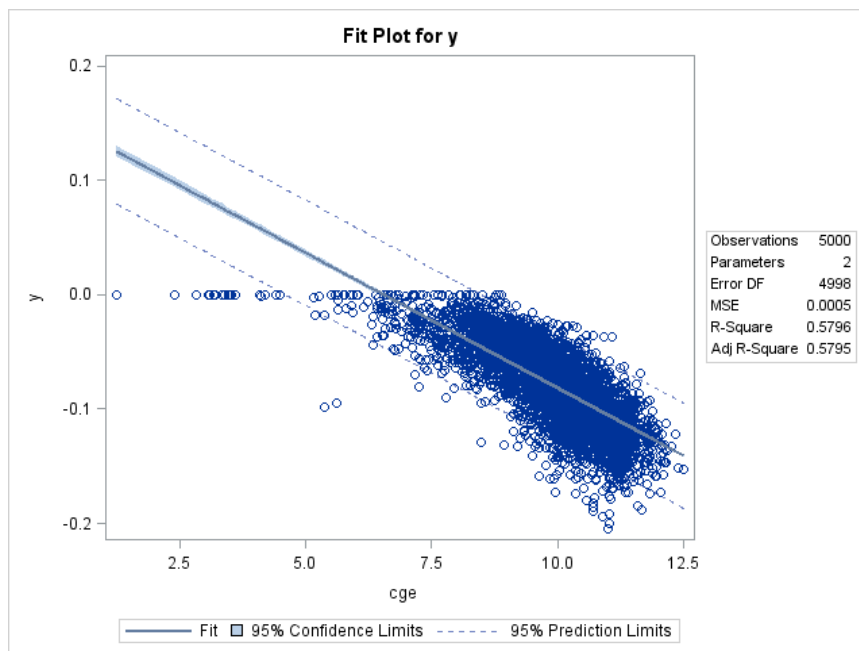
Metode	n	$\beta_1$	$R^2$	p-verdi
2	19056	-0.000924 $\pm$ 0.00000917	0,3477	<0,0001
4	5000	-0.02367 $\pm$ 0.00028519	0,5796	<0,0001
5	90490	0.00089540 $\pm$ 0.00000530	0,2399	<0,0001

Figur 4 viser regresjonsplot for Metode 2, hvor  $y = \ln(1-F)$  er regresset på fødselsår. Figuren viser en jevn fordeling av  $\ln(1-F)$  på fødselsår og punktene ligger godt fordelt rundt regresjonslinjen med få avvik utenfor 95% konfidenslinjen. Punktene som ligger som en tykk horisontal strek på nullpunktet er individer som mangler slektskapsinformasjon, hvor  $y=0$ .



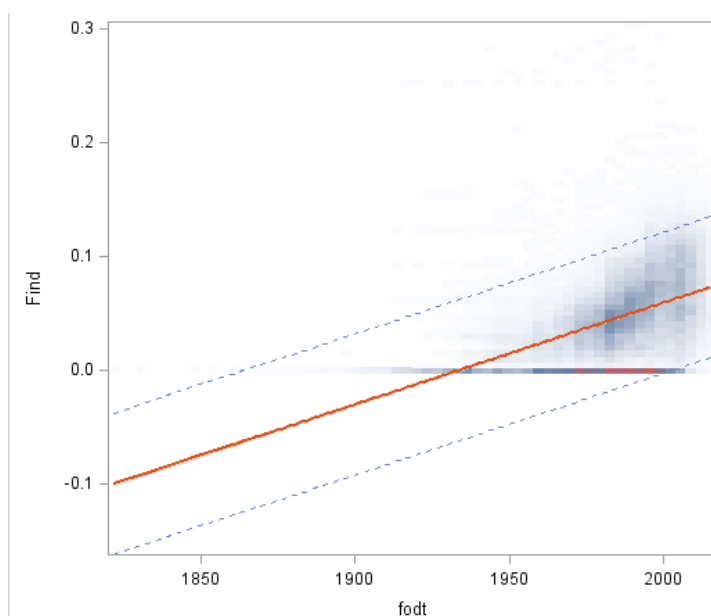
Figur 4. Viser regresjonsplot for Metode 2, hvor fødselsår er regresset på  $\ln(1-F)$ .

Figur 5 viser regresjonsplot for Metode 4 hvor det ble benyttet logregresjon og CGE ble regresset på  $\ln(1-F)$ . Plottet viser en bratt regresjonslinje, hvor punktsvermen har elliptisk form. Punktene blir tungt fordelt i nedre del av plottet hvor verdiene for CGE er høye, samtidig som F også har høye verdier.



Figur 5. Viser regresjonsplot for Metode 4, hvor CGE er regresset på  $\ln(1-F)$ .

Figur 6 viser regresjonsplott for Metode 5 hvor F er regresset på fødselsår. Plottet for denne metoden er relativt lik regresjonsplottet for Metode 2 som benytter logregresjon. Den har en normal regresjonslinje og punktene er jevnt fordelt.



Figur 6. Regresjonsplot for Metode 5, hvor F regresses på fødselsår.

### 5.5 Beregning av effektiv populasjonsstørrelse

Tabell 6 viser relativt stor variasjon ved beregning av effektiv populasjonsstørrelse mellom de fem metodene. Metode 1 skiller seg ut fra de andre metodene med en  $N_e$  på 128 dyr. Metode 3 og 4 er beregnet ved hjelp av CGE, hvor Metode 3 gir 2,75 ganger høyere effektiv populasjonsstørrelse enn metode 4. Metode 2 og 5 er regresset på fødselsår men har forskjellig antall individer, og det er i praksis ingen forskjell mellom metodene ved beregning av  $N_e$ .

Tabell 6. Resultatet fra beregningene av effektiv populasjonsstørrelse, estimert innavlsrate( $\Delta F_{gen}$ ) og standardfeil(SE) på de fem metodene. Metode 1 er basert på reelle paringer hvor de siste generasjonen er grunnlaget for den effektive populasjonsstørrelsen. Metode 2 er basert på logregresjon, hvor det er regresset på fødselsår. Metode 3 er basert på individets innavlskoeffisient og dens komplette generasjonsekivalent (CGE). Metode 4 er basert på logregresjon, regresset på komplett generasjonsekivalent. Metode 5 er basert på enkel lineær regresjon på fødselsår.

	n	$\Delta F_{gen}$	SE	$N_e$	Min $N_e$	Max $N_e$
Metode 1	5000	0,003893	5,96*	128	122,49	134,41
Metode 2	19054	0,000923	0,00000917	47	46,18	47,11
Metode 3	5000	0,008496	3,07*	59	55,78	61,92
Metode 4	5000	0,023392	0,00028519	21	21,12	21,63
Metode 5	90490	0,000895	0,00000530	47	47,16	47,72

\*beregnet  $SE(N_e)$  ved bruk av metoden til Gutiérrez et al.(2008) og modifisert av Cervantez et al. (2009).



## 6. Diskusjon

Tidligere er det blitt regnet innavlsrate ( $\Delta F$ ) og effektiv populasjonsstørrelse ( $N_e$ ) hos kaldblodstraveren (Olsen & Klemetsdal 2009), men dette er første gangen det er gjort på hele bestanden av varmblodstraver i Norge. Datamaterialet som er brukt i oppgaven inneholder alle registrerte dyr født frem til og med 2015. Det store datamaterialet gir grunnlag for en god pedigree, da mengden data er viktig for å få et godt resultat. Et stort datamateriale gir også grunnlag for en del feilkilder. Mange av individene i den originale slektskapsfilen hadde feil og mangler i slektskapsinformasjonen ble det gjort endringer for at disse skulle bli inkludert i datasettet. Endringene som ble gjort var tildeling av fiktive fødselsår, slette fiktive individer og slette dobbeltregistreringer av individer. Disse endringene kan føre til at slektskapsinformasjonen til individene predikerer høyere eller lavere verdier ved beregningene. Alle individene født i år 1700 fikk tildelt et fiktivt fødselsår basert på avkomsinformasjon, og antatt at foreldrene var født ett år før sine avkom (født-1), dette ble gjort for å rette opp individer hvor avkommene var født før foreldrene. I realiteten er det få oppdrettere/hesteeiere som bedekker en hest som ettåring, da med tanke på hoppens utvikling og egenskap til å være mor i en så ung alder (Vanderwall & McDonnell 2012). Mange av individene var registrert med fiktive dyr som foreldre, de fiktive foreldrene ble slettet da disse kan føre til ustabilitet i datasettet. De fiktive dyrene ble slettet på grunn av manglende slektskapsinformasjon, og ville blitt definert som foundere av populasjonen dersom de ble inkludert i slektskapsfilen. Inkluderingen av fiktive foreldre i slektskapsfilen ville kunne påvirket beregningene i Metodene 3 og 4 ved å forhøye verdiene av CGE.

### 6.1 Raseandeler

En studie gjort av Lykkjen et al. (2014) viser andel franske aner basert på 1217 hester, født 1988 og 2006/2007. Studien viser at andelen av franske aner hos varmblodstraveren i 1988 var relativt lav på kun 20 %, mens årgangen 2006/2007 var på 55 %. Den franske andelen hos varmblodstraverne har økt 35 % fra årgangene født 1988 til årgangen født 2006/2007. I denne oppgaven ble det beregnet franske raseandeler for individene i referansepopulasjonen ( $n=5000$ ), hvor 56 prosent av individene hadde franske aner (Figur 1). Det blir stadig mer populært å bedekke Amerikanske avlshopper med franske hingster og omvendt. Ser man på bedekningstallene for 2016 (Det Norske Travelskapet 2017) ser man at flere oppdrettere velger hingster som er krysning mellom amerikansk og fransk traver. Dette vil utgjøre en forskjell i beregnede innavlskoeffisienter for individer i senere

generasjoner, ved en større genetisk variasjon i avlsmaterialet. Sammenligner man Figur 1 og Figur 2 som beskriver de individuelle innavlskoeffisientene ser man hvordan de franske andelene preger de seks siste årgangene ved at innavlskoeffisientene synker.

Innavlskoeffisienten synker på bakgrunn av at den genetiske variasjonen har blitt bedre ettersom det før ble brukt kun amerikanske hingster på amerikanske hopper eller kun Fransk på Fransk i varmblodstraveravlen.

## **6.2 Genetisk fremgang og generasjonsintervall**

Den genetiske fremgangen( $\Delta G$ ) i en populasjon avhenger av sikkerheten i utvalget, mengden genetisk variasjon, seleksjonsintensiteten og generasjonsintervallet. For å få en størst mulig genetisk fremgang i en populasjon er det ønskelig med et kort generasjonsintervall, høy sikkerhet i utvalget, stor genetisk variasjon og høy seleksjonsintensitet (Klemetsdal 2014). Sammenhengen mellom disse faktorene er stor, et kort generasjonsintervall vil føre til at sikkerheten i utvalget blir lav og omvendt (Bourdon 2000). Unge dyr i avl vil ha mindre informasjon på de utvalgte egenskapene enn eldre dyr på grunn av avkomsinformasjon. Dette gjør at sikkerheten i utvalget blir mindre, men generasjonsintervallet kortere og raskere genetisk fremgang. I en artikkel av Decker (2014) hvor fokuset var hvordan man kunne endre generasjonsintervallet for å øke den genetiske fremgangen, viste han at en økning på 20% i sikkerheten i utvalget, utgjorde en 16,7% økning i  $\Delta G$ . En økning på 20% i seleksjonsintensitet utgjorde 20% økning i  $\Delta G$ , det viste seg at å redusere generasjonsintervallet fra 5 år til 4 år (20%) utgjorde en økning på den genetiske fremgangen med 25%. Dette viser at generasjonsintervall har en solid innvirkning på  $\Delta G$  og for å størst mulig fremgang bør man forsøke å finne en god balanse mellom sikkerheten i utvalget og generasjonsintervallet.

Det gjennomsnittlige generasjonsintervallet ligger på 11,60 per år for slektskaps-filen med 19056 individer og 11,76 per år for slektskaps-filen med 90490 individer.

Generasjonsintervallene i begge slektskaps-filene er relativt lange, og vil føre til en svakere genetisk fremgang i populasjonen. Årsaken til at generasjonsintervallene på varmblodstraveren er lange kan forklares med at det tar lang tid før hestene settes i avl på grunn av deres prestasjonsegenskaper. På varmblodstravere er den mest verdifulle egenskapen prestasjon, og det er dette oppdrettere ser på når de velger en hingst til sin avlshoppe. I tillegg til at det tar lang tid før hester settes i avl, tar det også lang tid før avkommene viser egen prestasjon og vil dermed forlenge tiden før hingsten får en god avlsverdi. Generasjonsintervallene i oppgaven viser symptomer på at de eldre hingstene

som har vist seg å ha god avlsverdi ved at avkommene har vist prestasjon, brukes i et lengre tidsrom enn de burde. Dette fører til lite variasjon i genmaterialet som vil øke sjansene for innavlsøkning i senere generasjoner, ettersom et langt generasjonsintervall vil gjøre at innavlsøkningen i populasjonen skjer over lengre tid. Det bør gjøres et mer effektivt avlsarbeid for å få hingstene inn i avlen tidligere og vesentlig forkorte generasjonsintervallet. Et alternativ er å bruke BLUP verdier mer aktivt i avlsarbeidet, eller genomisk seleksjon hvor man da kan få tilgang til dyrets avlsverdi allerede ved fødsel.

### ***6.3 Slektskapsdybde i datasettet***

I denne oppgaven er samtlige av metodene som blir brukt basert på slektskapsinformasjon og tilgjengeligheten av informasjon fra slektninger er viktig. En studie gjort av MacCluer et al. (1983) viste i sitt arbeid at ved undervurdering av distanserte generasjoner kan det føre til betydelig feilvurdering av innavl i en populasjon, i tillegg til at innavlskoeffisientene ( $F$ ) øker med dypere slektskapsinformasjon opptil 13 generasjoner. I denne oppgaven er gjennomsnittet for forfedre per individ (CGE) på 9,86 generasjoner, som tilsier at resultatene for de estimerte innavlskoeffisientene i denne oppgaven kan være underestimert i forhold til at slektskapsdybden i gjennomsnitt ligger på 9,86 og varierer fra 1,25 til 12,5 generasjoner individuelt i referansepopulasjonen.

I tråd med MacCluer et al. (1983) fant Pjontek et al. (2012) i sin studie at kvaliteten og mengden av slektskapsinformasjon påvirker de gjennomsnittlige innavlskoeffisientene. I denne oppgaven er slektskapsinformasjon i den totale slektskapsfilen på 19056 individer hentet tilbake i maksimalt 23 generasjoner. Tabell 2 viser slektskapsdybden i referansepopulasjonen ( $N=5000$ ), og viser at dybden er god i de 5 første generasjonene med andel kjente forfedre på 100-95 %. Slektskapsdybden synker gradvis frem til generasjon 10, og deretter faller den drastisk frem til generasjon 15 hvor andel kjente foreldre ligger på kun 2%. Årsaken til at andelen kjente foreldre synker skyldes manglende informasjon på individene jo lenger bakover i generasjonene en kommer. I figur 2 vises de gjennomsnittlige innavlskoeffisientene per år for 90490 individer. Figuren viser stor variasjon i innavlskoeffisientene på 1900-tallet, men blir mer stabile og øker gradvis frem til 2011 hvor innavlskoeffisientene synker. Variasjonen på 1900-tallet kan forklares med feil og mangler i den originale slektskapsfilen. Dette påvirker innavlskoeffisientene på grunn av for liten informasjon i slektskapet som kan føre til for lave estimerte innavlskoeffisienter. Kvaliteten på slektskapsfilen i denne oppgaven kan karakteriseres

som god på grunn av mange observasjoner, men ville vært bedre dersom registreringene på individene hadde vært komplette og uten feil og mangler.

#### **6.4 Forskjeller i metoder for beregning av $N_e$**

Metode 1 som baseres på Olsen et al. (2010) hvor  $\Delta F$  er beregnet ut fra innavlskoeffisientene til individene i referansepopulasjonen og deres foreldre. Metoden ga en  $\Delta F$  på 0,003893 og en effektiv populasjonsstørrelse på 128 dyr. Metode 1 gir en høy effektiv populasjonsstørrelse i forhold til de andre metodene. Metoden er en prospektiv metode, da den baserer seg på de seks siste årgangene i populasjonen og viser hvordan situasjonen er i populasjonen slik den er i dag basert på de siste seleksjonsavgjørelsene. Den høye effektive populasjonsstørrelsen kan forklares med en nedgang i innavlskoeffisientene for de siste årgangene, Figur 2 viser hvordan de gjennomsnittlige innavlskoeffisientene til populasjonen bøyer av de siste årgangene. Dette fører da til en lav  $\Delta F$  og kan forklares med at interessen for krysning av Amerikansk og Fransk travet har blitt mer populært de seneste årene.

Metode 2 som baserer seg på metoden til Pérez-Enciso (1995) hvor det benyttes logregresjon som regresser på fødselsår gir  $\Delta F$  på 0,000923 og den effektive populasjonsstørrelse på 47 dyr. Figur 4 viser et fitplot fra regresjonsanalysen av metode 2, den viser en normal regresjonslinje og punktene er jevnt fordelt rundt regresjonslinjen med få punkter utenfor konfidenslinjene. Metode 2 og 4 som begge benytter logregresjon viser store forskjeller ved regress på fødselsår og CGE, det skiller 26 dyr i den effektive populasjonsstørrelsen. Når man sammenligner Figur 4 og 5 kan virke som at regresjon på fødselsår er en mer robust metode for beregning av  $\Delta F$  enn regresjon på CGE. I doktorgradsavhandlingen til Hillestad (2015) ble det benyttet  $\ln(1-F)$  regresset på fødselsår og  $\ln(1-F)$  regresset på CGE med data på Norsk Rødt Fe. Arbeidet viste at  $\ln(1-F)$  regresset på CGE ga en høyere  $R^2$  enn  $\ln(1-F)$  regresset på fødselsår som indikerte at variasjonen ble godt forklart ved bruk av CGE i denne modellen. I denne oppgaven ga også  $\ln(1-F)$  regresset på CGE en høyere  $R^2$  enn  $\ln(1-F)$  regresset på fødselsår, men ser man på Figur 5 som viser fitplot av regresjonsanalysen hvor CGE er benyttet kan man se den bratte regresjonslinjen og den elliptiske formen på punktsvermen. I tillegg har modellen flere punkter utenfor konfidenslinjene som tyder på at modellen for regresjon på CGE ikke er en god modell for denne populasjonen.

To av metodene som ble testet i denne oppgaven benytter CGE i beregningene for innavlsraten. Metode 4 som baseres på Hillestad et al. (2015) benytter logregresjon der

$\ln(1-F)$  regresses på CGE og Metode 3 som baseres på Gutierrez et al. (2009) og benytter individuell CGE og individuelle innavlskoeffisienter. Metodene gir stor variasjon i beregningene for  $\Delta F$ , selv om begge benytter CGE.  $\Delta F$  i Metode 3 er 2,8 ganger lavere enn  $\Delta F$  i Metode 4, dette gjør at forskjellen i den effektive populasjonsstørrelse ved bruk av disse metodene er stor. At metodene har stor variasjon i  $\Delta F$  kan skyldes mange grunner. En av grunnene kan være de komplette generasjonsekvivalentene, da disse kun er gjennomsnittlige verdier og vil kunne påvirke metodene i forskjellige retninger. I denne oppgaven ble det beregnet individuell CGE for individene i referansepopulasjonen ( $n=5000$ ) med gjennomsnitt på 9,86 generasjoner. Beregningene av CGE viser lite variasjon som kan skyldes at enkelte individer i referansepopulasjonen har flere kjente foundere, som gir et dypere slektskap på enkelte individer. Dette fører til høy innavlskoeffisienter hos individer med en høy CGE verdi. I Figur 5 kan man se hvordan overrepresentasjonen av høye innavlskoeffisienter ( $F$ ) ved høy CGE danner en ellipseform i punktene. I Metode 4 fører dette til at regresjonslinjen blir brattere og forårsaker overestimering av  $\Delta F$ . Metode 3 har i motsetning til Metode 4 gitt lav  $\Delta F$ . Som sagt tidligere er det lite variasjon i CGE, og ved bruk av Metode 3 vil det kunne antas at verdien på CGE kan være for liten i forhold til det den burde vært da metoden bruker  $^{CGE_i-1}\sqrt{(1 - F_i)}$  som kan gi en mindre verdi og dermed underestimere innavlsraten ( $\Delta F$ ) som gir en høyere effektiv populasjonsstørrelse (Gunnar Klemetsdal pers. anm.).

Metode 5 som baseres på enkel lineær regresjon hvor det regresses på fødselsår på 90490 individer ga  $\Delta F$  på 0,000895 og effektiv populasjonsstørrelse på 47 dyr. Denne regresjonsmetoden gir ikke store forskjeller i beregningene av  $N_e$  sammenlignet med Metode 2 som benytter logregresjon på fødselsår. Figur 6 viser fitplottet til Metode 5 og viser en jevn fordeling av punktene rundt regresjonslinjen, men har en del uteliggere utenfor konfidenslinjen som gjør at graden av feilestimatene blir større. Enkel lineær regresjon er en alminnelig metode for å estimere  $\Delta F$ , men på grunn av lave innavlskoeffisienter i responsvariabelen vil logregresjon være en bedre metode å bruke fordi logregresjonen minsker graden av feilestimatene i modellen (Tuft 2000).

Ved bruk av Metode 2 og 5 kan generasjonsintervallet skape en underestimering av  $N_e$ . Generasjonsintervallet påvirker  $\Delta F$  i stor grad og fører til økning i  $\Delta F$  ved multiplikasjon. Det vil si at i disse metodene blir  $\Delta F$  noe høyere på grunn av generasjonsintervallet og man bør derfor prøve å unngå beregninger hvor generasjonsintervallet inkluderes.

Formålet med denne oppgaven var å teste metoder for beregning av innavlsrate og effektiv populasjonsstørrelse ved bruk av fem forskjellige metoder. Resultatene fra metodene vises i Tabell 5. Tabellen viser stor variasjon i  $\Delta F$  og dermed store forskjeller i den effektive populasjonsstørrelsen i metodene. Ved regresjonsmetodene ville det vært mulig å avklare hvor gode modellene er ved bruk av  $R^2$ , men på grunn av ulike antall observasjoner i modellene vil det ikke være hensiktsmessig. Det mest hensiktsmessige vil være å se på fitplottene for regresjonsmodellene. I denne oppgaven vil Metode 2 hvor logregresjon på fødselsår være den beste metoden for beregning av  $N_e$  da fitplottet for metoden viser en jevn fordeling av punktene med få avvik utenfor konfidenslinjene. Ved bruk av Metode 3 som beregnes basert på individuelle innavlskoeffisienter og individuell CGE, hvor den effektive populasjonsstørrelsen ligger på et brukbart nivå ( $N_e=59$ ) vil man være kritisk til metoden på grunn av beregningen av CGE i denne populasjonen.

## 7. Konklusjon

Det er viktig å vite hvilken metode som ligger til grunn for beregningene av  $N_e$  da denne vil påvirke hva som gjenspeiles i resultatene. Ut ifra beregningene i denne oppgaven bør man være skeptisk til bruken av CGE i populasjonen for varmbloodstraver, fordi det er lite variasjonen i CGE-verdiene kan det føre til over- og underestimering av  $\Delta F$  i Metode 3 og 4. I tillegg bør det unngås bruk av generasjonsintervall i beregningene for  $N_e$  da høye generasjonsintervall vil påvirke  $\Delta F$  og vil kunne underestimere  $N_e$ .

Logregresjon på fødselsår er en bedre metode å bruke enn enkel lineær regresjon på fødselsår i dette datasettet, fordi feilmarginsverdiene i datasettet er betydelig redusert i logregresjonen enn i den enkel lineære regresjonen. Metode 1 som baserer seg på de seks siste generasjonene, vil sammen med Metode 2 (logregresjon på fødselsår) kunne være gode metoder å bruke for å vurdere genetisk variasjon i populasjonen ut fra en praktisk vinkling.

## Referanser

- Boichard, D., Maignel, L., Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genetic Selection Evolution*, 29: 5-23.
- Bourdon, R. M. (2000). *Understanding Animal Breeding, 2nd Edition*. Factors Affecting the Rate of Genetic Change: Pearson.
- Cervantes, I., Goyache, F., Molina, A., Valera, M. & Gutiérrez, J. P. (2008). Application of individual increase in inbreeding to estimate realized effective sizes from real pedigrees. *J. Anim. Breed. Genet.*
- Cervantes, I., Goyache, F., Molina, A., Valera, M. & Gutierrez, J. P. (2011). Estimation of effective population size from the rate of coancestry in pedigreed populations. *J Anim Breed Genet*, 128 (1): 56-63.
- Decker, J. (2014). Decreasing Generation Interval to Increase Genetic Progress. Tilgjengelig fra: <http://articles.extension.org/pages/72665/decreasing-generation-interval-to-increase-genetic-progress> (lest 06.Jan.2017).
- Det Norske Travelskapet. (2015). Norsk avlsplan for varmlods travhester.
- Det Norske Travelskapet. (2017). *Bedekninger 2016*. Travsport.no. Tilgjengelig fra: <http://www.travsport.no/Nyhetsarkiv/2017/Februar/Bedekninger-2016/> (lest 22. Feb).
- Econ. (2009). Hest i Norge. *Econ-rapport nr. 2009-001*. 51 s.
- Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Small populations: I. Change of gene frequency under simplified conditions. II. Less simplified conditions: Longman Group.
- FAO. (1998). Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans- Management of Small Populations at Risk.
- Gutierrez, J. P., Cervantes, I., Molina, A., Valera, M. & Goyache, F. (2008). Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genet Sel Evol*, 40 (4): 359-378.
- Gutierrez, J. P., Cervantes, I. & Goyache, F. (2009). Improving the estimation of realized effective population sizes in farm animals. *J Anim Breed Genet*, 126 (4): 327-32.
- Hillestad, B. (2015). *Inbreeding determined by the amount of homozygous regions in the genome*: Norwegian University of Life Science, Department of Animal and Aquacultural Sciences. 138 s.



- Hillestad, B., Woolliams, J. A., Meuwissen, T., Våge, D. I. & Klemetsdal, G. (2015). Estimating Rate of Inbreeding and Effective Population Size using Genomic Data in Norwegian Red Cattle.
- Klemetsdal, G. (2014). Forelesning HFA200. .
- Lykkjen, S., Olsen, H. F., Dolvik, N. I., Grondahl, A. M., Roed, K. H. & Klemetsdal, G. (2014). Heritability estimates of tarsocrural osteochondrosis and palmar/plantar first phalanx osteochondral fragments in Standardbred trotters. *Equine Vet J*, 46 (1): 32-7.
- Løvås, G. G. (2010). *Statistikk for Universiteter og Høgskoler, 2.utgave*. Enkel lineær regresjon: Universitetsforlaget AS.
- MacCluer, J. W., Boyce, A. J., Dyke, B., Weitkamp, L. R., Pfenning, D. W. & Parsons, C. J. (1983). Inbreeding and pedigree structure in Standardbred horses. *The Journal of Heredity*, 74: 394-399.
- Microsoft Excel. (2016). Microsoft AS.
- Moureaux, S., Verrier, É., Ricard, A. & Mériaux, J. C. (1996). Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms. *Genet Sel Evol*, 28: 83-102.
- Norsk Hestesenter. (2003). Avlsplan for varmbloodstraver. *Starum*.
- Olsen, H. F. & Klemetsdal, G. (2009). Innavlsutvikling hos kaldbloodstraver.
- Olsen, H. F., Klemetsdal, G., Ruane, J. & Helfjord, T. (2010). Pedigree structure and genetic variation in the two endangered Norwegian horse breeds: Døle and Nordland/Lyngen. *Acta Agriculturae Scandinavica, Section A - Animal Science.*, 60: 13-22.
- Pérez-Enciso, M. (1995). Use of the uncertain relationship matrix to compute effective population size. *Anim. Breed. Genet.*, 112: 327-332.
- Pjontek, J., Kadlecik, O., Kasarda, R. & Horny, M. (2012). Pedigree analysis in four Slovak endangered horse breeds. *Czech J. Anim.Sci.*, 57: 54-64.
- Samsonstuen, S. (2015). *Effekt av innavl på prestasjonsegenskaper hos varmbloodstraver i Norge*: Norges Miljø- og Biovitenskaplige Universitet, Institutt for Husdyr- og Akvakulturvitenskap.
- SAS. (2013). *SAS/CONNECT® 9.4 User's Guide*, Second Edition: Cary, NC: SAS Institute Inc.
- Tufte, P. E. (2000). En intuitiv innføring i logistisk regresjon. *Arbeidsnotat nr 8.*: s 1-82.

- Vanderwall, D. & McDonnell, S. (2012). *Appropriate Breeding Age*. The Horse -your guide to equine health care. Tilgjengelig fra:  
<http://www.thehorse.com/articles/28515/appropriate-breeding-age> (lest 3. Jan).
- Vangen, O. (2016a). *Hest*. Store Norske Leksikon. Tilgjengelig fra: <https://snl.no/hest> (lest 7. November).
- Vangen, O. (2016b). *Varmblodstraver*. Store norske leksikon. Tilgjengelig fra:  
<https://snl.no/varmblodstraver> (lest 15 November).
- Vik, J. & Farstad, M. (2012). Hest, hestehold og fôring: Status for hesteholdet i Norge. Norsk senter for bygdeforskning: Norges Forskningsråd og Stiftelsen Svensk Hästforskning. 56 s.
- Wright, S. (1931). Evolution In Mendelian Populations. *Genetics*, 16: 97-159.



Norges miljø- og biovitenskapelig universitet  
Noregs miljø- og biovitenskapelige universitet  
Norwegian University of Life Sciences

Postboks 5003  
NO-1432 Ås  
Norway