



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2018 30 stp

Fakultet for Biovitenskap

Hovedveileder: Gunnar Klemetsdal

Predikerte avlsverdier for egenskaper på det Norske Landsvinet funnet ved bruk av prinsipal komponent analyse

Predicted breeding values on traits from the
Norwegian Landrace found using principal
component analysis

Ellen Maud Langmoen Olsen

Husdyrvitenskap

Instituttet for husdyr- og akvakulturvitenskap

Forord

Denne oppgaven markerer avslutningen på en 2-årig mastergrad i husdyrvitenskap, ved fakultet biovitenskap ved Norges miljø- og biovitenskapelige universitet. Allerede i det første året på universitet ble det klart at hovedinteressen min var innenfor avl og genetikk. Det er mye spennende som skjer innen avl på gris, derfor var det et naturlig valg for meg.

Denne masteren er en fortsettelsen på bacheloroppgaven jeg skrev våren 2016. Som handlet om å gjøre en variabelreduksjon på egenskapene i avlsmålet til Landsvin og Duroc. Masteroppgaven har beveget seg litt vekk fra den opprinnelige ideen, men det var bacheloroppgaven som la grunnlaget.

Takk til Norsvin for at jeg har fått lov til å bruke deres datamateriale i denne masteren.

Jeg vil rette en stor takke til min hovedveileder professor Gunnar Klemtesdal som har bistått med mange ideer og gode innspill. En takk til min biveileder Ina Margrete Andersen-Ranberg som har lært meg alt jeg kan om både SAS og DMU, samt mye mer! Takk til biveileder Jørgen Kongsro som var personen bak ideen til masteren og hjelp under hele prosessen. Til slutt takk til Tale Marie Karlsson-Drangsholt som både har laget datasettet som er brukt i denne masteroppgaven og bidratt med faglig kompetanse underveis.

Jeg vil også takke min samboer Nobuki som både har hjulpet meg med Matlab programmering og lest oppgaven flere ganger. Takk til min Astrid Langmoen Olsen, Margit Langmoen Olesn og Gina Langmoen Olsen som har lest og rettet denne oppgaven. Til slutt men ikke minst takk til verdens søteste lille hund, som har gjort at denne prosessen har vært mye enklere for meg.

Ås, mai 2018

Ellen Maud Langmoen Olsen

Sammendrag

Avlsarbeidet til det Norske Landsvinet blir driftet av Norsvin. I dag er det 27 egenskaper med i avlsmålet, men det brukes 49 egenskaper når en beregner avlsverdier. Disse egenskapene blir delt inn i grupper hvor det blir beregnet varianser og avlsverdier på en gruppe av egenskaper om gangen. Dette gjør at kovariansene mellom gruppene blir 0. Det er derfor behov for en metode som kan nyttiggjøre seg av alle egenskapene Norsvin har, og som samtidig kan beregne varianskomponenter og avlsverdier på et stort antall egenskaper i en utregning. Formålet med denne oppgaven var å bruke data fra Norsvin til å gjennomføre en prinsipalkomponent analyse på et utvalg av egenskaper registrert på Landsvin. Et mål var å utnytte prinsipalkomponentene som har egenverdi høyere enn 1,0 til å beregne varianser og avlsverdier i DMU, for så å tilbakekonvertere de til opprinnelig skala. Et annet mål var å sammenligne disse nye avlsverdiene med tilsvarende avlsverdier regnet ut med samme gruppering som det Norsvin bruker i dag, på samme data.

I dette forsøket har det blitt brukt data fra Norsvin, hvor det etter begrensningene ble brukt 1687 dyr med fenotyper på 41 egenskaper, og 6992 dyr i slektskapsmatrisen. Det ble gjort en PCA på de 41 egenskapene, hvor det ble 19 PC med en egenverdi høyere enn 1,0. De 19 PC ble brukt i DMU for å beregne varianser og avlsverdier. Avlsverdiene til de 19 PC ble tilbakekonvertert ved hjelp av ladninger laget da det ble gjort en PCA. De tilbakekonverterte avlsverdiene ble sammenlignet med avlsverdiene funnet ved bruk av den konvensjonelle metoden som Norsvin benytter i dag. Denne sammenligningen var basert på kryssvalidering.

Oppgaven viser at det via PCA er mulig å beregne avlsverdier for hele 41 egenskaper om gangen. PCA-tilnærmingen viser ingen klar sammenheng mellom eksteriøregenskapene og produksjonsegenskapene. Kryssvalideringen viser at noen egenskaper predikeres bedre via PCA enn den konvensjonelle tilnærmingen. Det blir bedre prediksjoner for noen av egenskapene siden informasjonen brukes til å bestemme færre varianskomponenter via PCA, enn ved den konvensjonelle metoden. I tillegg blir det kovarianser på tvers av alle egenskapene, hvor dette ikke er tilfelle for gruppekjøringene. Det ser ut til at det ekstra informasjon å hente ved å bruke PCA, derfor burde denne metoden videre testes.

Abstract

Norsvin runs the breeding program of the Norwegian Landrace. As of today, they use 49 traits when breeding values are predicted, but only 27 of these are used for breeding goals. The traits are divided into groups, where each group separately are used to estimate variances and breeding values. The covariances between the groups are therefore 0. Because of this, there is a need for a method which can make use of all the traits Norsvin has data on, while also being able to estimate variance components and breeding values on a larger number of traits. The purpose of this thesis was to use data from Norsvin to run a principal component analysis on a selection of traits, registered on the Landrace. One goal was to use the calculated principal components which had an eigenvalue above 1.0, to estimate variances and breeding values in DMU, before converting them back to the original scale. A second goal was to compare these new values with equivalent breeding values, which was calculated using the same data, and the same grouping that Norsvin currently uses.

The data used in this thesis was provided by Norsvin. The pedigree file used contained 6992 animals, of which 1687 of them had phenotypes on 41 traits. When a PCA was executed on these traits, there were 19 PC with an eigenvalue above 1.0. They were then used in DMU to estimate variances and breeding values on PCs. These new values, together with loadings given by the PCA, were converted back to breeding values on the 41 original traits. The breeding values found on the traits were compared with the breeding values found using the conventional method that Norsvin uses today. This comparison was based on cross-validation.

This thesis shows that with the use of PCA it is possible to estimate breeding values on 41 traits without having to divide the traits into different groups for estimation. Using PCA shows no clear correlation between exterior traits and production traits. Cross-validation shows that some traits are better predicted with PCA rather than the conventional method. The reason for this is that the information is used to decide fewer variance components through PCA, compared to conventional method. With PCA, one will also get covariances between the traits, while this is not the case when the traits are divided into groups. It seems like it is more information that can be gained using PCA, and the method should, therefore, be more tested.

Innhold

Forord.....	I
Sammendrag	III
Abstract.....	V
1 Innledning.....	1
2 Teori.....	3
2.1 Landsvinet	3
2.2 Dagens avlsarbeid	3
2.3 DMU.....	8
2.4 Prinsipal Komponent Analyse.....	8
3 Material og metode.....	11
3.1 Datamateriale og utvalgsriterier	11
3.2 Konvensjonell metode.....	13
3.2.1 Modellene for den konvensjonelle metoden	15
3.3 Prinsipal Komponent Analyse.....	19
3.3.1 Modellen for prinsipalkomponentene i DMU.....	19
3.4 Tilbakekonvertering av prinsipalkomponentene	20
3.5 Korrelasjonsanalyse	21
3.6 Kryssvalidering	21
4 Resultater	23
4.1 Korrelasjonen mellom konvensjonell-EBV og PC-EBV	23
4.2 Kryssvalidering	26
4.3 Genetiske korrelasjoner mellom PC.....	29
5 Diskusjon.....	31
6 Konklusjon.....	35
Kilder	37

1 Innledning

Det norske Landsvinet er en av de beste morrasene i verden på egenskaper som tilvekst, fôrforbruk og moregensskaper. Dette kommer av at Norsvin har hatt et godt organisert avlsarbeid og at alvsbesetningene gjør en god jobb. I dag blir Landsvinet (både levende individer og sæd) eksportert til store deler av verden siden den er robust og presterer bra i mange ulike miljø samt at den har en veldig god helse (Andersen-Ranberg, 2016).

For å få til avlsfremgang er det viktig at det er mange og gode registreringer på hver enkelt egenskap. Ved Norsvin sin teststasjon, Delta, blir det testet 3500 råner hvert år, hvor ca. halvparten av dem er Landsvin. For alle rånene som kommer til teststasjon registreres det informasjon på egenskaper som tilvekst og fôrforbruk daglig på hvert enkelt individ. I tillegg blir det gjort en eksteriørbedømming, og til slutt blir de CT-skannet. CT gir fenotyper av høy kvalitet på kroppssammensetningen til avlskandidaten (Martinsen et al., 2016). Siden det bare er råner på teststasjon, kommer informasjonen om moregenskapene fra besetningene. De viktigste besetningene på Landsvin er foredlingsbesetningene. Det er disse som driver med renrasert Landsvin. Dette er de samme besetningene som leverer råner til teststasjonen (*Foredlingsbesetninger*, 2012). Bonden registrer moregenskaper som antall fødte, kullvekt og antall avvente grisunger (Andersen-Ranberg, 2016). Dette gjør at tanter, mødre og søstre til avlsrånene har fenotyper for disse egenskapene.

På Landsvinet fenotypes det mer enn 60 ulike egenskaper. Det vil si at Norsvin har mye informasjon, men de benytter bare en del av dette. I avlsmålet på Landsvin er det nå 27 egenskaper, men når det beregnes avlsverdier, brukes 49 egenskaper. De 49 egenskapene blir delt opp i ulike grupper, hvor det blir beregnet varianser og avlsverdier for hver enkelt gruppe via DMU. Grunnen til at disse egenskapene blir delt opp i grupper er fordi DMU ikke greier å estimere varianskomponenter og avlsverdier på alle disse egenskapene samtidig i én utregning (Andersen-Ranberg, 2018). Kovarianser er viktig å ha med i avlsarbeidet, da disse gir informasjon om hvordan to eller flere egenskaper samsvarer. Det er derfor behov for en metode som kan nyttiggjøre seg av alle egenskapene Norsvin har, og som samtidig greier å beregne varianskomponenter og avlsverdier på et så stort antall egenskaper i en utregning.

Formålet med denne oppgaven var å bruke data fra Norsvin til å gjennomføre en prinsipal komponent analyse på et utvalg av egenskaper registrert på Landsvin. Et mål var å utnytte prinsipalkomponentene som har egenverdi høyere enn 1,0 til å beregne varianser og avlsverdier i DMU, for så å tilbakekonvertere dem til opprinnelig skala. Et annet mål var å sammenligne disse nye avlsverdiene med tilsvarende avlsverdier regnet ut med samme gruppering som det Norsvin bruker i dag, på samme datamateriale. Denne sammenligningen var basert på kryssvalidering.

2 Teori

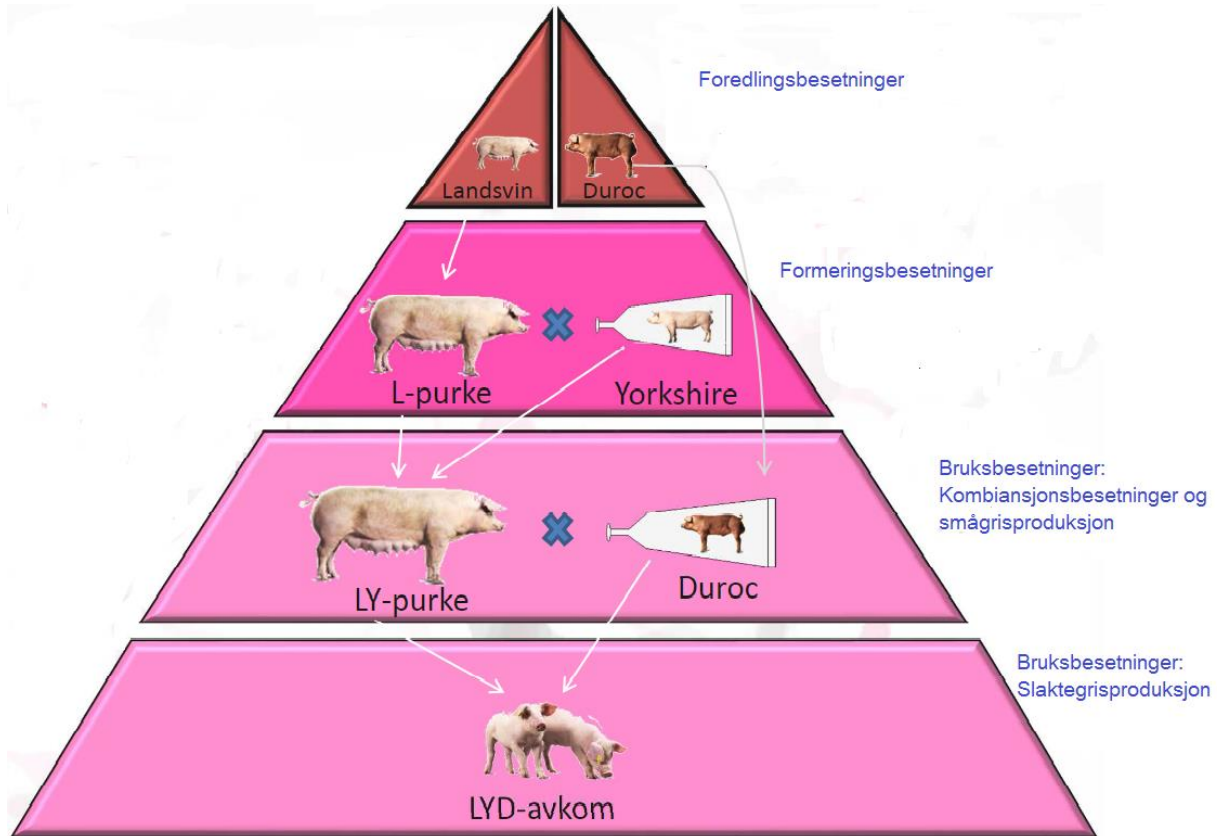
2.1 Landsvinet

Landsvinet er en hvit rase, med lang kropp, relativt lite hode og hengende ører (Vangen et al., 1994). Avlsarbeidet på Landsvinet startet på slutten av 1950-tallet, med fokus på tilvekst, fôrforbruk og mindre spekktykkelse (Gjerlaug-Enger, 2011). Landsvinet har de siste 20 årene blitt avlet med tanke på å kombinere slaktegrisens effektivitet og morslinjens produktivitet, og i dag er Landsvinet ansett som en av de beste morrasene i tillegg til at den er god på produksjonsegenskaper. Derfor blir genene til Landsvinet eksportert over hele verden (Andersen-Ranberg, 2016).

2.2 Dagens avlsarbeid

Norsvin er en avlsorganisasjon som er eid av bøndene og som driver avlsarbeid på både Landsvin og Duroc. Svineproduksjonen i Norge baserer seg på et krysningsopplegg, derfor er det forskjell på hvilke egenskaper som vektlegges i de ulike rasene. Landsvinet er en morrase, og avles derfor på moregenskaper. Landsvinet blir krysset med rasen Yorkshire fra Nederland, som blir distribuert av Topigs Norsvin. Topigs Norsvin er i dag den nest største leverandøren av svinegener i verden (Martinsen et al., 2016). Duroc blir avlet på som en farrase, det vil si mest for produksjonsegenskaper og slaktekvalitet (Gjerlaug-Enger, 2011).

Siden svineproduksjonen i Norge baserer seg på krysningsavl, er det nødvendig med en del besetninger som driver reinavl på rasene. Avlssystemet er derfor bygget opp i form av en pyramide (se figur 1).



Figur 1: Avlssystemet i Norge (Andersen-Ranberg, 2016).

På toppen av pyramiden er foredlingsbesetningene. Disse besetningene rekrutterer dyr til råneteststasjonen, og selger landsvinpurker til formeringsbesetninger. Foredlingsbesetningene står for en rekke registreringer som er viktige på egenskaper som levendefødte, antall avvente, helse og hvor mange dager det tar fra avvenning til purka igjen er bedekt. Dette er viktige informasjon som en ikke får målt på teststasjonen (*Foredlingsbesetninger*, 2012). Neste ledd nedover er formeringsbesetningene. Disse kjøper Landsvin fra foredlingsbesetninger og inseminerer disse med Yorkshire-sæd (som blir distribuert fra Topigs Norsvin), for å produsere mordyrkrysningen TN70 (*Formeringsbesetninger*, 2013). TN70 purkene blir solgt videre til bruksbesetninger. TN70 purka blir også eksportert (levende individer), og har blitt en viktig morraskryssning internasjonalt, siden krysningen lager gode og robuste mordyr. Bruksbesetningene er den siste delen av pyramiden, men også den største, for det er disse som er hovedprodusentene av svinekjøtt i Norge. Bruksbesetningene kan deles inn i tre driftsformer;

kombinasjonsbesetninger, smågrisprodusenter og slaktegrisbesetninger. Smågrisprodusentene krysser TN70 purka med en råne, gjerne Duroc. Smågrisen blir gjerne solgt videre til slaktegrisbesetninger, som fører opp smågrisen til den er slakteklar. Tilsvarende produserer kombinasjonsbesetningen smågris og beholder de til de er slakteklare. Slaktegrisen er altså en kryssning av Landsvin, Yorkshire og Duroc, selv om Landsvin og Hampshire også blir brukt som farraser (*Bruksbesetninger*, 2013; Andersen-Ranberg, 2016). Selv om det er foredlingsbesetningene som er det viktigste grunnlaget for videre avl, er registreringene (gjennom Ingris) fra alle de andre besetningene nedover i pyramiden også viktige. Ingris er her et viktig verktøy som kan brukes av alle bønder. Her registreres viktig informasjon om purkene, smågrisene og slaktegrisene som er viktig å ta med seg videre i avlsarbeidet (Martinsen et al., 2016). Bonden registrerer mye ulik informasjon som senere kan bli hentet ut. Ved mer informasjon vil de predikerte avlsverdiene bli sikrere (*Ingris*, 2016).

Norsvin har avlskonsulenter som reiser rundt til foredlingsbesetningene for å gjøre unnggrismålinger. Dette er en felttest hvor avlskonsulenten måler blant annet vekt, spekktykkelse, muskel og bedømmer eksteriør på unnggris når grisene er ca. 150 dager gamle (Andersen-Ranberg, 2016).

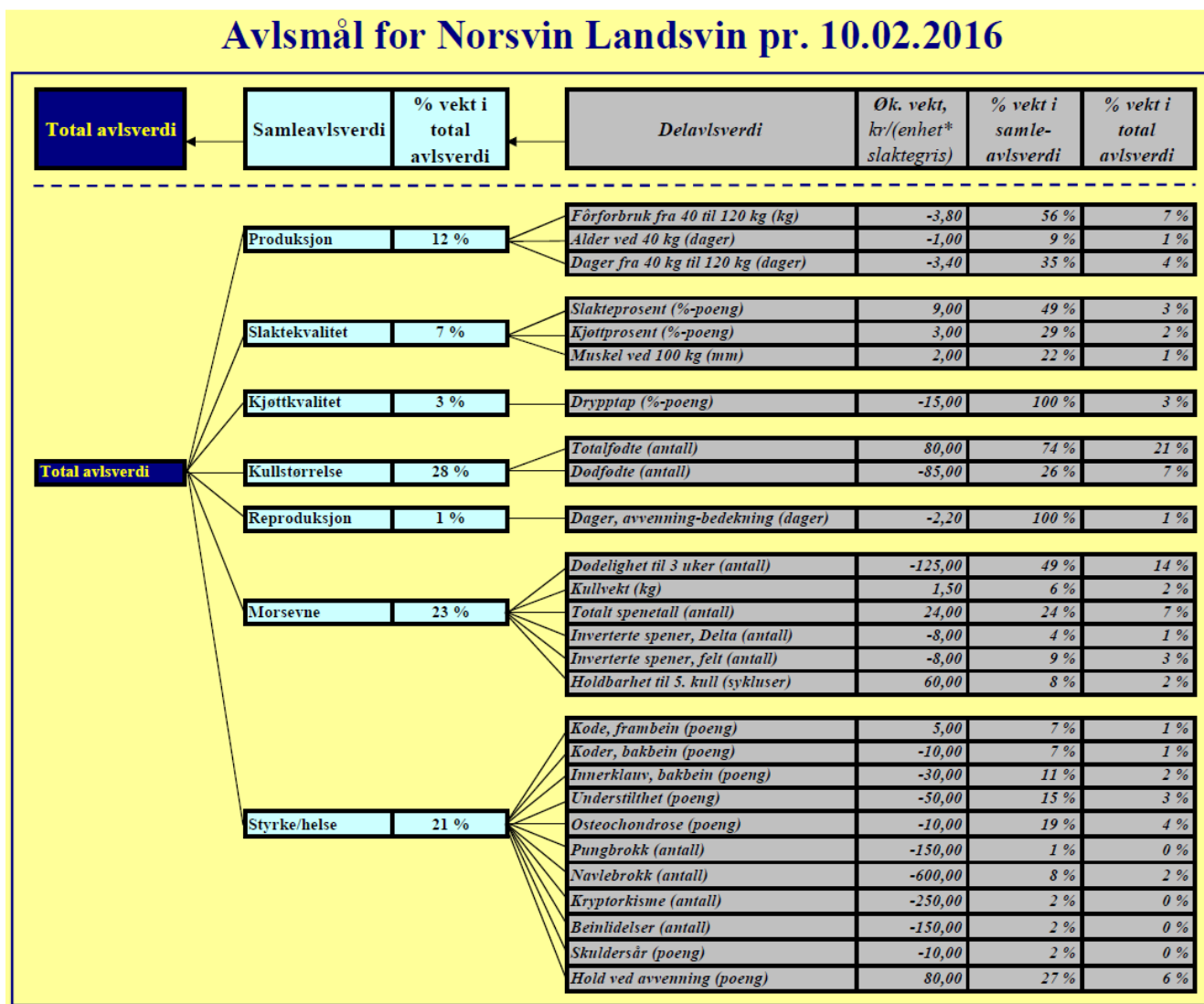
Det er 27 foredlingsbesetninger med Landsvin, med til sammen ca. 2 100 purker. Det blir ca. født 4 400 kull i året av rasen Landsvin i Norge. Fra disse kullene blir det utvalgt 1 200 råner som blir fraktet til teststasjon Delta. Disse blir utvalgt på grunnlag av deres foreldres estimerte avlsverdi, hvor da Norsvin gir beskjed til foredlingsbesetningene om at de ønsker 1-2 råner fra de beste kullene. Produsenten velger da de beste rånene i kullet med minst 14 spener, god tilvekst, tilfredsstillende eksteriør samt at de ellers overholder kravene Norsvin stiller. Av de 1 200 rånene blir 130 senere seminrånere, for så til slutt å bli 50 eliterånere (Martinsen et al., 2016).

I tillegg til informasjonen som kommer fra besetningene gjennom Ingris og unnggrismåling, blir avlskandidatene (1 200 råner) testet på Delta. Rånene blir fraktet til Delta når de er mellom 30-35 kg, og fra første dag blir tilvekst og fôrforbruket registrert ved hjelp av «FIRE» stasjoner i bingene. Disse gir individuelle målinger på hvor mye grisen til enhver tid veier og spiser. Når grisen veier 120 kg blir testen ved Delta avsluttet. Det blir gjort en grundig eksteriørbedømming og grisen blir CT-skannet (Martinsen et al., 2016).

Datatomografi (CT) ble installert i 2008 på Norsvin sin teststasjon, og i dag brukes den rutinemessig i avlsarbeidet (Gjerlaug-Enger, 2011). CT fungerer slik at røntgenstrålene blir sendt gjennom kroppen og absorbert i vev med høy tetthet, som skjelett med mye mineraler. I vev med lav tetthet, som fett, vil derimot røntgenstrålene gå igjennom nesten uabsorbert. Det er mengden absorbert/uabsorbert som danner basis for røntgenbildene, i ulike gråskalaer. Skjelett som har høy tetthet vil være farget helt hvitt på bildet, mens vev med lav tetthet vil ha en mye mørkere gråtone (Brekke & Kolbenstvedt, 2014). CT gjør at en kan beregne blant annet kjøttprosent og slakteprosent på selve avlskandidatene (Kongsro, 2015), slik at en kan finne fenotypene uten å måtte benytte halvsøskentest for disse egenskapene. CT blir også brukt for å lage et digitalt anatomisk kart av kroppen til grisen; et Atlas. Atlaset vil gjøre det mulig å segmentere grisen, automatisk skille mellom skjelettet, muskler, innvoller og kunne dele kroppen inn i stykkeningsdeler. Slik kan en få mer detaljert informasjon om grisen enn det vanlig CT greier å gi oss i dag (Kongsro, 2017). Atlaset gir fenotyper for stykkeningsdelene på avlsrånene, og utfordringen nå er å benytte denne kunnskapen videre i avl. Det brukes «maskinlæring» til fenotyper etter CT-skanning (Gangsei & Kongsro, 2015). Utfordringen nå er å ta disse nye fenotypene i bruk i avlsarbeidet.

Kjøttkvalitetsegenskaper som drypptap og farge er pr i dag vanskelig å finne på selve avlskandidaten. Derfor blir det registrert kjøttkvalitetsegenskaper på de rånene som slaktes (ikke selektert til seminråne) når de er ferdig testet på Delta (Gjerlaug-Enger, 2011). De slaktede rånene vil gjennom sitt slektskap til seleksjonskandidaten gi informasjon om kjøttkvalitetsegenskapene. I denne sammenhengen blir det tatt en DNA-test for å beregne slektskapet mellom dyrene. Før antok en at alle fullsøsken seg imellom delte et like stort slektskap, og det samme for halvsøsken. Det betydde at alle fullsøsken delte 50 % av de samme genene, og at halvsøsken delte 25 % av de samme genene. I dag vet man at fullsøsken kan være mer eller mindre i slekt med hverandre enn 50%. Ved å ta en DNA-test kan en finne enkeltindividets innbyrdes slektskap og benytte denne informasjonen i beregningene av avlsverdier, altså genomisk slektskap. Dette betyr at informasjon fra slektninger som har en høyere andel slektskap vil få en større betydning i predikeringen av avlsverdeier (Andersen-Ranberg, 2016).

Avlsmålet til Landsvin i dag har med de 27 egenskapene en kan se i figur 2. De fleste av egenskapene her blir fenotypet på Delta, med unntak av moregenskaper som blir testet ute i besetninger og kjøttkvalitet som blir testet etter slaktning. Avlsverdiene blir beregnet automatisk hver natt for alle 49 egenskaper (Andersen-Ranberg, 2018).



Figur 2: Avlsmålet for Landsvin (Andersen-Ranberg, 2016).

2.3 DMU

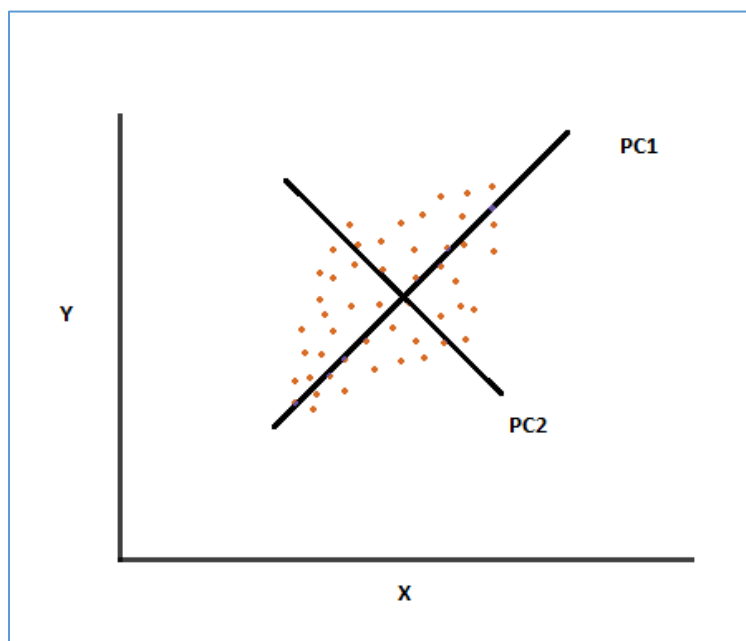
Avlsverdier beregnes med DMU multivariat. Da vil hver egenskap gi informasjon om alle de andre egenskapene (Lynch & Walsh, 1998). Fordelen med DMU er dette programmet også kan beregne varianskomponenter. Dette betyr at en bare trenger et program for å beregne begge deler. Med mange egenskaper i beregningen av varianskomponenter blir det svært mange parametere som skal bestemmes (spesielt i kompliserte modeller) som ofte fører til at REML-algoritmen (REstricted Maximum Likelihood) ikke konvergerer og en må derfor begrense antallet egenskaper som blir tatt med.

Norsvin løser denne svakheten til DMU ved at ulike grupper av egenskaper blir kjørt sammen. I stedet for å ha med alle de 49 egenskapene i en utregning blir de delt inn i ulike grupper som produksjon, slaktekvalitet, morsevne, osv. (se figur 2). Derfor blir varianskomponenter og avlsverdeier estimert gruppevis. Avlsverdiene for hver egenskap blir standardisert for så å bli multiplisert med hver sin økonomisk vekt (som Norsvin har gitt dem). De økonomiske vektene kan ses i figur 2. Avlsverdiene for hver enkelt egenskap blir også vektet sammen til en total avlsverdi (Andersen-Ranberg, 2018). Når en estimerer avlsverdier gruppevis så tas det hensyn til korrelasjonene mellom egenskapene innenfor hver enkelt gruppe, men det tas ikke hensyn til korrelasjonene mellom egenskapene på tvers av gruppene.

2.4 Prinsipal Komponent Analyse

Prinsipal komponent analyse (PCA) er en statistisk prosedyre som bruker en ortogonal transformasjon for å konvertere et sett med variabler, eller egenskaper, som kan være korrelerte, til et sett med verdier av lineært ukorrelerte variabler kalt prinsipalkomponenter (PC). Formålet med PCA kan være å finne underliggende mønster i datasettet. Den første PC vil beskrive mest av variasjonen i dataen, hvor de påfølgende PC har den nest meste variasjonen osv. I figur 3 er dette fremstilt med to PC som står vinkelrett på hverandre (PC1 og PC2). En tredje PC vil ligge på en vinkelrett retning til både PC1 og PC2. Til slutt vil det være et PC-system som består av en rekke ortogonale PC, som alle ligger langs en retning med maks varians i avtagende rekkefølge. Siden alle PC er gjensidige ortogonale så vil de være ukorrelerte med hverandre. Det kan aldri være flere PC enn det originale antallet variabler. PC1 vil alltid være den som beskriver det meste av variansen. Ved dannelsen av stadig flere PC vil variansen som forklares bli stadig

mindre, hvor de siste PC vil ligge i en retning hvor det lenger nesten ikke er noe spredning i observasjonene, som i praksis betyr at det ikke lenger er flere underliggende strukturer igjen i datasettet (Esbensen, 2000). Ved bruk av PCA får en et innsyn i de underliggende strukturene i dataen som kan gi økt forståelse hvis PC tolkes riktig. Via PCA vil gjerne antallet PC ofte være redusert relativt til det opprinnelige antallet variabler. Dette vil kunne redusere datakraften som er nødvendig for å beregne varianskomponenter og avlsverdier, siden det blir et redusert antall varianskomponenter som trenger å bli estimert.



Figur 3: En illustrasjon av to uavhengige Prinsipalkomponenter som står ortogonalt på hverandre. PC1 beskriver den største andelen av variansen i datasettet, og PC2 nest mest.

Antall PC som en velger å beholde kan baseres på størrelsen av egenverdien. I figur 3 beskriver egenverdien til PC1 hvor mye punktene er spredt ut fra «linja» til den første PC (Dallas, 2013). Alle PC har hver sin egenverdi. Antallet PC en velger å bruke videre kan bestemmes på flere ulike måter. En metode er å velge de PC som har en egenverdi høyere enn 1,0. En grunn til å velge denne metoden er at egenverdi mindre enn 1 gir mindre informasjon enn hva som blir gitt av en enkelt variabel (NCSS, 2018).

Ved beregning av PCA blir det laget, i tillegg til PC, noe som heter ladninger. Denne gir oss informasjon om sammenhengen mellom de originale verdiene og de nye PC-verdiene. Dette er en matrise som kan bli brukt for å transformere PC-verdiene tilbake til de originale variablene (Esbensen, 2000).

Det er fenotypene en lager PC av. Selv om det ikke er en fenotypisk korrelasjon mellom PC, kan det fortsatt være en genetisk korrelasjon mellom dem. Dette betyr at ved bruk av PCA vil en kunne utnytte de genetiske korrelasjonene på tvers av PC, som da vil ha potensiale for å gjøre dem bedre enn de nåværende gruppevise utregningene, hvor de genetiske korrelasjonene settes til 0 på tvers av gruppene.

Det er viktig å finne ut hvordan en PCA-tilnærming gjør det relativt til den vanlige gruppeutregningen som blir brukt for å beregne avlsverdier. En metode er å beregne avlsverdier ved bruk av begge tilnærminger for så å sammenligne dem. Dette kan gjøres ved å se på korrelasjonen mellom avlsverdiene til de originale egenskapene og avlsverdiene funnet ved bruk av PC. Alternativt kan en benytte kryssvalidering. Det finnes flere måter å kryssvalidere på, men en metode er å ekskludere noen individer fra det opprinnelige datasettet. Siden de ekskluderte dyrene er med i slektskapsfilen så vil det bli predikert avlsverdier på dem. På denne måten kan det bli beregnet en korrelasjon mellom de nye predikerte avlsverdiene opp mot de ekskluderte dyrenes fenotyper. Dette gir en korrelasjon mellom avlsverdiene og fenotypene for hver av tilnærmingene, og den tilnærmingen som får den høyeste korrelasjonen mellom avlsverdi og fenotype kan ses på som best (Andonov et al., 2013).

3 Material og metode

3.1 Datamateriale og utvalgskriterier

Datamaterialet i denne oppgaven er hentet fra Norsvin sin database, Pigfo, på Landsvin. Det meste av informasjonen på egenskapene kommer fra råneteststasjonen, Delta, med unntak av egenskapen vekt ved 21 dager (V21) som kommer fra foredlingsbesetningene. I denne oppgaven var det tilgang på 1830 dyr med fenotyper på 41 ulike egenskaper. Slektskapsfila besto av 6992 dyr, hvor denne inneholder alle de 1830 dyrene med observasjoner og forfedrene deres. Det krevdes i denne oppgaven at alle rånene har fullstendig informasjon på alle egenskapene med kun én observasjon på hver egenskap.

Egenskapene i dataen ble delt inn i tre gruppeegenskaper: eksteriør, Atlas og produksjon. Se tabell 1, 2 og 3. I dette forsøket ble det ikke tatt med moregenskaper siden purkene ikke blir CT-skannet, dermed får de ikke informasjon om Atlasegenskapene.

På det tidspunktet dette datamaterialet ble laget hadde ikke Norsvin flere enn 1830 dyr med informasjon om Atlasegenskapene, og siden kravet var at alle dyrene skulle ha informasjon på alle egenskapene, så begrenset en seg her til disse 1830 dyrene.

Den konvensjonelle metoden å beregne avlsverdier på blir forklart først i denne oppgaven, for det var reduksjonene gjort i denne delen som la grunnlaget for dataen som senere ble brukt i PCA.

Tabell 1: Eksteriøregenskaper som er med i dataen, med en beskrivelse av hver enkelt egenskap.

Eksteriøregenskaper:	Beskrivelse av egenskapene
Fb_kode	Frambein kode
Bb_kode	Bakbein kode
Fb_kne	Forbein kne
Bb_hase	Bakbein hase
Bb_innerklauv	Bakbein innerklauv
Rygg_dokk	Rygg dokk
Understilt	Understilt
Svingbakpart	Svingende bakpart

Tabell 2: Egenskaper som bestemmes ved bruk av datatomografi (CT) som er med i dataen, og en beskrivelse av hver enkelt egenskap.

CT egenskaper:	Beskrivelse av egenskapen
Kjøttpro	Kjøttprosenten angir andelen av slaktet som er kjøtt (CT)
Slaktepro	Slakteprosent bestemmes ved CT, altså hvor stor del som består av bein, muskler, fett osv. (korrigert til dager)
Shoulder_p	Hvor mange prosent skulderen utgjør av hele slaktet (Atlas)
Shoulder	Hvor mange kg av grisen som er skulder (Atlas)
Shoulder_meat	Hvor stor andel av skulderen som er kjøtt (kjøttprosent i skulderen) (målt ved bruk av Atlas)
Belly_p	Hvor mange prosent buken utgjør av hele slaktet (Atlas)
Belly	Hvor mange kg av grisen som er buken (Atlas)
Belly_meat	Hvor stor andel av buken er kjøtt (kjøttprosent) (målt ved bruk av Atlas)
Loin_p	Hvor mange prosent kammen utgjør av hele slaktet (Atlas)
Loin	Hvor mange kg av grisen som er kammen (Atlas)
Loin_meat	Hvor stor andel av kammen som er kjøtt (Atlas)
Ham_p	Hvor mange prosent skinka utgjør av hele slaktet (Atlas)
Ham	Hvor mange kg av grisen som er skinka (Atlas)
Ham_meat	Hvor stor andel av skinka som er kjøtt (Atlas)
Tenderloin_p	Hvor mange prosent indrefiletten utgjør av hele slaktet (Atlas)
Guts	Hvor mange kg innvollene utgjør av hele slaktet (Atlas)
Guts_meat	Hvor stor andel av innvollene som er kjøtt (Atlas)
Testicles	Hvor mange kg testiklene utgjør av hele slaktet til grisen (Atlas)
Lmp	Kjøttprosent (Lean Meat Percentage) (Atlas)
Yield	Slakteprosent (Atlas)
Sum_ocd	Den totale summen av osteokondrose (CT)

Tabell 3: Produksjonsegenskaper som er med i dataen, med en beskrivelse av hver enkelt egenskap.

Produksjonsegenskaper:	Beskrivelse av egenskapen
V21	Hvor mye grisen veier ved tre uker, målt i besetning
M100	Muskeltykkelse i mm målt ved siste ribbein korrigert til levendevekt på 100kg (CT)
S100	Ryggspektykkelse i mm målt ved siste ribbein korrigert til levendevekt på 100kg (CT)
Muscle	Muskeltykkelse i mm målt ved siste ribbein (ikke korrigert)
Backfat	Ryggspekk i mm målt ved siste ribbein (CT)
A40	Alder (dager) ved 40 kg.
D40_80	Dager brukt på å vokse fra 40kg til 80 kg (levendevekt)
D80_120	Dager brukt på å vokse fra 80 til 120kg (levendevekt)
D40_120	Dager brukt på å vokse fra 40 til 120kg (levendevekt)
F40_80	Fôrmengde brukt på å vokse fra 40 til 80 kg (levendevekt)
F80_120	Fôrmengde brukt på å vokse fra 80 til 120kg (levendevekt)
F40_120	Fôrmengde brukt på å vokse fra 40 til 120 kg (levendevekt)

3.2 Konvensjonell metode

For å estimere varianser og predikere avlsverdier på så mange av de 41 egenskapene som mulig, ble modellen gjort enklest mulig og inneholdt bare de tilfeldige effektene; *individ* og *feil*.

Følgende tilfeldige effekter ble sett bort ifra: bingen dyret har stått i på Delta (binge), gruppenummer (group_no), kullet dyret er født i (kull) og ID til mora til dyret (mor). For at fjerning av kulleffekten ikke skulle påvirke variansestimeringen for mye ble det beholdt bare et individ fra hvert kull. Dette reduserte datasettet ned til 1687 dyr. Alle de faste effektene ble beholdt, samt alle regresjonene (disse blir beskrevet senere).

Selv etter at de tilfeldige effektene ble fjernet, var det fortsatt ikke mulig å ha bare en kjøring med de 41 egenskapene. Derfor ble egenskapene gruppert i fire, som går fram av tabell 4. For at det ikke skulle bli hele fem grupper, ble tre egenskaper fjernet; dager brukt for å vokse fra 80 til 120 kg (d80_120), fôr brukt for å vokse fra 80 til 120kg (f80_120) og skulder (antall kg av slaktet som er skulder).

Tabell 4: Egenskaper som er i de ulike gruppekjøringene i DMU ved bruk av den vanlige metoden.

Gruppe 1: Multivariat 1 egenskaper	Gruppe 2: Multivariat 2 egenskaper
V21	Muscle
M100	Backfat
S100	D40_80
A40	F40_80
D40_120	Kjottpro
f40_120	Slaktepro
Shoulder_p	Belly
Shoulder_meat	Loin
Belly_p	Ham
Belly_meat	Guts
Sum_ocd	Guts_meat
Gruppe 3: Atlasegenskaper	Gruppe 4: Eksteriøregenskaper
Loin_p	Fb_kode
Loin_meat	Bb_kode
Ham_p	Fb_kne
Ham_meat	Bb_hase
Tenderloin_p	Bb_innerklauv
Testikler	Rygg_dokk
Lmp	Understilt
Yield	Svingbakpart

3.2.1 Modellene for den konvensjonelle metoden

Modellene brukt i denne oppgaven tar utgangspunkt i de modellene Norsvin bruker i sin avlsverdiberegning. Modellene ble kjørt i DMU, versjon, release 5.2 (Madsen & Jensen, 2013). Modellene tar utgangspunkt i den generelle mixed modellen (se ligning 1). Nedenfor blir modellene brukt for de konvensjonelle egenskapene, forklart for hver enkelt gruppe.

Modellene for de 19 PC tar utgangspunkt i den generelle mixed modellen (se ligning 1). Nedenfor blir modellene i de ulike gruppene forklart.

Den generelle mixed modellen (Lynch & Walsh, 1998):

$$Y = X\beta + Zu + e \quad [1]$$

hvor Y er observasjonen på egenskapen, β er vektoren for de faste effektene, u er vektoren for de tilfeldige effektene og e er feilen (miljøfaktoren). X og Z er matriser som kobler observasjonene til faste (X-matrisa) og tilfeldige effekter (Z-matrisa).

Gruppe 1

Gruppe 1 er en kjøring som kombinerer flere forskjellige egenskaper. I DMU blir det beregnet varianser og avlsverdier på alle disse egenskapene i en kjøring, men det blir brukt fire forskjellige modeller. Felles for alle egenskapene er at det er den samme tilfeldige effektene som inngår i modellen; *individ* og *feil*. Hvilke faste effekter og regresjoner som blir brukt, varierer. Modellene blir beskrevet nedenfor.

Egenskapene v21, m100, s100 og a40 har modellen:

$$Y_{jgfk} = medlaar_j + m_kullnr_g + fmnd_f + individ_i + b_1levf_k + b_2levf_{2+} + e_{jgfk}$$

hvor Y er observasjonen av egenskapene (V21, m100, s100 og a40). $Medlaar_j$ er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3). $Fmnd_f$ er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). $Individ$ er en tilfeldig effekt hvor alle har et individuelt nummer. $Levf_k$ er en regresjoner som beskriver antall levendefødte i kullet, det er 23

nivåer ($k = 1, 2, 3, \dots, 23$), og $levf_{2l}$ antall levendefødte i kullet kvadrert, 23 nivåer ($l = 1, 4, 9, \dots, 592$). Til slutt er det e_{jfgkl} som er en tilfeldig effekt av feil.

Egenskapene d40_120 og f40_120 har modellen:

$$Y_{jgfhik} = medlaar_j + m_kullnr_g + fmnd_f + avdeling_h + individ_i + b_1levf_k + e_{jgfhik}$$

hvor Y er observasjonen av egenskapene (d40_120 og f40_120). $Medlaar_j$ er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3). $Fmnd_f$ er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). $Avdeling$ er en fast effekt som sier hvilken avdeling grisen har stått i på Delta, det er 16 nivåer ($h = 1, 2, 3, \dots, 16$). $Individ$ er en tilfeldig effekt. $Levf_k$ er en regresjoner som beskriver antall levendefødte i kullet har, 23 nivåer ($k = 1, 2, 3, \dots, 23$). Til slutt er det e_{jgfhik} som er en tilfeldig effekt av feil.

Egenskapene Shoulder_p, Shoulder_meat, Belly_p og Belly_meat har modellen:

$$Y_{jgfhip} = medlaar_j + m_kullnr_g + fmnd_f + avdeling_h + individ_i + b_1lw_p + e_{jgfhip}$$

hvor Y er observasjonen av egenskapene (Shoulder_p, Shoulder_meat, Belly_p og Belly_meat). $Medlaar_j$ er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3 kull). $Fmnd_f$ er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). $Avdeling$ er en fast effekt som sier hvilken avdeling grisen har stått i på Delta, det er 16 nivåer ($h = 1, 2, 3, \dots, 16$ avd.). $Individ$ er en tilfeldig effekt. Lw_p er en regresjon som beskriver levendevekten til grisen, den har 1389 nivåer ($p = 108,8, 111,5, \dots, 140,7$ kg). Til slutt er det e_{jgfhip} som er en tilfeldig effekt av feil.

Egenskapen sum_ocd:

$$Y_{gsaip} = m_kullnr_g + oc-dommer_s + HY_a + individ_i + b_1lw_p + e_{gsaip}$$

hvor Y er observasjonen av egenskapen (sum_ocr). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3 kull). Oc_dommer_s sier hvilken person det er som bedømmer mengden osteokondrose, det er 2 nivåer ($s =$ dommer 39 og 41). HY er hvilken besetning og år individet kommer fra, og er en fast effekt den har 82 nivåer ($a = 6015, 9214, \dots, n$). $Individ$ er en tilfeldig effekt. Lw_p er en regresjon som beskriver levendevekten til grisen, den har 1389 nivåer ($p = 108,8, 111,5, \dots, 140,7$ kg). Til slutt er det e_{gsaip} som er en tilfeldig effekt av feil.

Gruppe 2

Gruppe 2 er en også en multivariat kjøring. I denne kjøringen blir det bukt tre forskjellige modeller hvor felles for alle egenskapene er at det er en tilfeldige effektene; *individ* og *feil*. De faste effektene og regresjonene som blir brukt varierer.

Egenskapene Muscle og Backfat har modellen:

$$Y_{jgfikl} = medlaar_j + m_kullnr_g + fmnd_f + individ_i + b_1levf_k + b_2levf2_l + e_{jifgkl}$$

hvor Y er observasjonen av egenskapene (Muscle og Backfat). $Medlaar_j$ er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3). $Fmnd_f$ er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). $Individ$ er en tilfeldig effekt hvor alle har et individuelt nummer. $Levf_k$ er en regresjoner som beskriver antall levendefødte i kullet har, 23 nivåer ($k = 1, 2, 3, \dots, 23$), og $levf2_l$ antall levendefødte i kullet kvadrert, 23 nivåer ($l = 1, 4, 9, \dots, 592$). Til slutt er det e_{jifgkl} som er en tilfeldig effekt av feil.

Egenskapene d40_80, f40_80, kjottpro og slaktepro har modellen:

$$Y_{jgfhik} = medlaar_j + m_kullnr_g + fmnd_f + avdeling_h + individ_i + b_1levf_k + e_{jgfhik}$$

hvor Y er observasjonen av egenskapene (d40_80, f40_80, kjottpro og slaktepro). $Medlaar_j$ er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3). $Fmnd_f$ er fødselsmåned og er

en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). *Avdeling* er en fast effekt som sier hvilken avdeling grisen har stått i på Delta, det er 16 nivåer ($h = 1, 2, 3, \dots, 16$). *Individ* er en tilfeldig effekt. $Levf_k$ er en regresjoner som beskriver antall levendefødte i kullet, det er 23 nivåer ($k = 1, 2, 3, \dots, 23$). Til slutt er det e_{jgfhik} som er en tilfeldig effekt av feil.

Egenskapene Belly, Loin, Ham, Guts og Guts_meat:

$$Y_{jgfhip} = medlaar_j + m_kullnr_g + fmnd_f + avdeling_h + individ_i + b_1lw_p + e_{jgfhip}$$

hvor Y er observasjonen av egenskapene (Belly, Loin, Ham, Guts og Guts_meat). *Medlaar_j* er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). *M_kullnr_i* er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3 kull). *Fmnd_f* er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). *Avdeling* er en fast effekt som sier hvilken avdeling grisen har stått i på Delta, det er 16 nivåer ($h = 1, 2, 3, \dots, 16$ avd.). *Individ* er en tilfeldig effekt. Lw_p er en regresjon som beskriver levendevekten til grisen, den har 1389 nivåer ($p = 108,8, 111,5, \dots, 140,7$ kg). Til slutt er det e_{jgfhip} som er en tilfeldig effekt av feil.

Gruppe 3

I denne kjøringen har alle egenskapene den samme modellen. Her er modellen for disse egenskapene:

$$Y_{jgfhip} = medlaar_j + m_kullnr_g + fmnd_f + avdeling_h + individ_i + b_1lw_p + e_{jgfhip}$$

hvor Y er observasjonen av egenskapene (Loin_p, Loin_meat, Ham_p, Ham_meat, Tenderloin_p, Testikler, LMP og Yield). *Medlaar_j* er en fast effekt som beskriver medlemsnummeret til grisen og hvilket år den er født, 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). *M_kullnr_i* er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3 kull). *Fmnd_f* er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). *Avdeling* er en fast effekt som sier hvilken avdeling grisen har stått i på Delta, det er 16 nivåer ($h = 1, 2, 3, \dots, 16$ avd.). *Individ* er en tilfeldig effekt. Lw_p er en regresjon som beskriver levendevekten til grisen, den har 1389 nivåer ($p = 108,8, 111,5, \dots, 140,7$ kg). Til slutt er det e_{jgfhip} som er en tilfeldig effekt av feil.

Gruppe 4

I denne kjøringen har også alle egenskapene den samme modellen. Her er modellen for disse egenskapene:

$$Y_{wgavi} = officer2_w + m_kullnr_g + HY_a + YS_v + individ_i + e_{wgavi}$$

hvor Y er observasjonen av egenskapene (Fb_kode, Bb_kode Fb_kne, Bb_hase, Bb_innerkluav, Rygg_dokk, Understilt og Svingbakpart). $Officer2_w$ er en fast effekt som beskriver hvilken operatør det er som har gjort målingen, det er 2 nivåer ($w = \text{operatør 2 og 41}$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2 \text{ eller } 3 \text{ kull}$). HY er hvilken besetning og år individet kommer fra, og er en fast effekt denn har 82 nivåer ($a = 6015, 9214, \dots, n$). YS_v er en fast effekt som sier hvilket år og hvilken sesong det er 15 nivåer ($v = 20144, 20145, \dots, 20166$). $Individ$ er en tilfeldig effekt. Til slutt er det e_{wgavi} som er en tilfeldig effekt av feil.

3.3 Prinsipal Komponent Analyse

Prinsipal komponent analysen ble kjørt i Matlab. Dette ble gjort på 1687 individer, siden dette er antallet som ble brukt i den konvensjonelle metoden. Rånene hadde informasjon på alle de 41 egenskapene, men med kun én observasjon på hver egenskap.

PCA modellen kan matematisk bli beskrevet som (Dagnachew et al., 2012):

$$w = (I_n \otimes P) * t + \mathcal{E}, \quad [2]$$

hvor w er en vektor av alle de 41 egenskapene. t er en vektor av score (prinsipalkomponenter). I_n er identitetsmatrisen av størrelsen n , hvor n står for antall egenskaper som er tatt med, og P er matrisen for ladninger. \otimes er et Kronecker produkt, og er \mathcal{E} en vektor av feil.

Etter at prinsipal komponent analysen hadde blitt kjørt i Matlab, ble score verdiene for PC med egenverdi høyere enn 1,0 hentet ut og kjørt med DMU.

3.3.1 Modellen for prinsipalkomponentene i DMU

En valgte å ha med alle de faste effektene og regresjonene som blir brukt i modellene til den konvensjonelle metoden. I dette forsøket ble det estimert varianser og predikert avlsverdier for alle 19 PC.

Modellene brukt for alle PC i DMU:

$$Y_{javgfhwsiklp} = medlaar_j + HY_a + YS_v + m_kullnr_g + fmnd_f + avdeling_h + officer2_w + oc-dommer_s + individ_i + b_1levf_k + b_2levf2_l + b_3lw_p + e_{javgfhwsiklp}$$

hvor Y er observasjonen av egenskapene (alle de 19 PC). $Medlaar_j$ er en fast effekt som beskriver medlemsnummeret til boden hvor grisen er født og hvilket år den er født, det er 79 nivåer ($j = 4015, 4016, 7416, \dots, n$). HY er hvilken besetning (hvor grisen har blidt målt, kan være en annen besetning enn hvor den ble født) og år individet kommer fra, og er en fast effekt den har 82 nivåer ($a = 6015, 9214, \dots, n$). YS_v er en fast effekt som sier hvilket år og hvilken sesong det er, 15 nivåer ($v = 20144, 20145, \dots, 20166$). M_kullnr_i er en fast effekt som beskriver hvilket kullnummer det er for mora til individet som blir testet, denne har 3 nivåer ($i=1, 2$ eller 3 kull). $Fmnd_f$ er fødselsmåned og er en fast effekt denne har 12 nivåer ($f = 1, 2, 3, \dots, 12$ mnd.). $Avdeling$ er en fast effekt som sier hvilken avdeling grisen har stått i på Delta, det er 16 nivåer ($h = 1, 2, 3, \dots, 16$ avd.). $Officer2_w$ er en fast effekt som beskriver hvilken operatør det er som har gjort målingen, det er 2 nivåer ($w =$ operatør 2 og 41). $Oc-dommer_s$ sier hvilken person det er som bedømmer mengden osteokondrose, det er 2 nivåer ($s =$ dommer 39 og 41). $Individ$ er en tilfeldig effekt. $Levf_k$ er en regresjoner som beskriver antall levendefødte i kullet har, 23 nivåer ($k = 1, 2, 3, \dots, 23$), og $levf2_l$ antall levendefødte i kullet kvadrert, 23 nivåer ($l = 1, 4, 9, \dots, 592$). Lw_p er en regresjon som beskriver levendevekten til grisen, den har 1389 nivåer ($p = 108,8, 111,5, \dots, 140,7$ kg). Til slutt er det $e_{javgfhwsiklp}$ som er en tilfeldig effekt av feil.

3.4 Tilbakekonvertering av prinsipalkomponentene

De predikerte avlsverdiene på de 19 PC ble hentet ut fra DMU, og tatt inn i Matlab. Dette ble gjort for at avlsverdiene skulle bli tilbakekonvertert ved hjelp av ladningene som tidligere hadde blitt laget.

En formel for tilbakekonvertering er gitt av Dagnachew et al. (2012):

$$\tilde{\mu}_w = (I_n \otimes P) \cdot \tilde{\mu}_t \quad [3]$$

hvor $\tilde{\mu}_w$ er de predikerte avlsverdiene etter de har blitt konvertert tilbake. I_n er identitetsmatrisen av størrelsen n , hvor n står for antall egenskaper som er tatt med, og P er matrisen for ladninger.

\otimes er et Kronecker produkt, som er en metode for å beregne to matriser sammen, og $\tilde{\mu}_t$ er de predikerte avlsverdiene for PC, funnet ved hjelp av DMU.

Tilbakekonvertering av avlsverdiene for de 19 PC ga nye avlsverdier for alle de opprinnelige 41 egenskapene. På denne måten ble avlsverdiene for de 19 PC brukt for å finne avlsverdiene for alle de 41 egenskapene.

3.5 Korrelasjonsanalyse

Etter avlsverdiberegning ved bruk av 1687 dyr, ble alle de predikerte avlsverdiene hentet inn i SAS. Dette gjaldt avlsverdiene fra de fire konvensjonelle gruppeutregningene og de tilbakekonverterte avlsverdiene fra PC funnet i DMU. Alle de fem filene ble slått sammen i SAS, for deretter å beregne en korrelasjon mellom de konvensjonelle avlsverdiene og PC vlaverdiene via SAS versjon 9.4.

3.6 Kryssvalidering

For å finne ut hvilken metode som predikerte de mest korrekte avlsverdiene, så ble det gjennomført en kryssvalidering. Dette ble gjort ved at det ble laget et nytt datamateriale som ekskluderte 296 individer som hadde fenotypiske observasjoner på alle de 41 egenskapene, fra det originale datasettet. Disse 296 dyrene ble beholdt i et eksternt datasett, for at datasettet senere skulle bli brukt for å se hvilken metode som hadde høyest korrelasjon mellom fenotype og avlsverdi. Det nye datasettet som hadde blitt redusert ned til 1391 dyr ble brukt til å beregne nye avlsverdier, ved bruk av begge metoder. Varianskomponentene som ble funnet i de tidligere utregningene ble beholdt, da disse er mer korrekte, siden de ble funnet på et større datasett. Avlsverdiene fra de fire ulike gruppene ble hentet ut med SAS og slått sammen med de 300 ekskluderte dyrene som hadde blitt fjernet tidligere. Deretter ble det beregnet korrelasjoner mellom de nye avlsverdiene og fenotypene på de 296 dyrene som ikke var med i den siste avlsverdiberegningen. Det samme ble gjort for PC avlsverdiene. Korrelasjonsanalysen ble gjort i SAS (SAS, 2013).

4 Resultater

Tabell 5 viser resultatene fra PCA gjennomført på 41 egenskaper. I denne tabellen er det tatt med 19 PC, siden det var så mange som hadde en egenverdi høyere enn 1,0. De 19 PC beskrev så mye som 99% av den totale variansen i dataen.

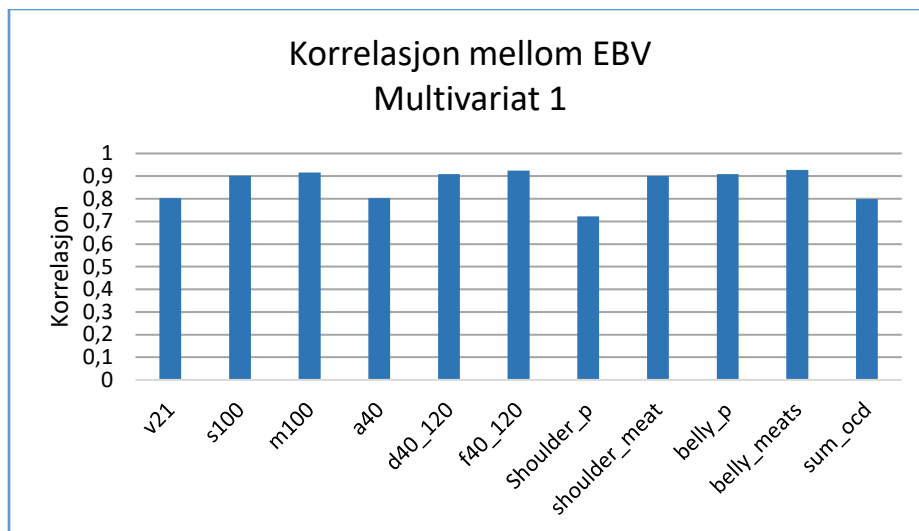
Tabell 5: Egenverdien og variansen beskrevet for hver av de 19 PC som ble tatt med i dette forsøket.

PC	Egenverdi	Varians
1	212,28	47,79
2	69,02	15,54
3	36,66	8,25
4	28,62	6,44
5	24,72	5,57
6	18,86	4,25
7	9,51	2,14
8	8,68	1,95
9	7,42	1,67
10	5,10	1,15
11	3,88	0,87
12	3,12	0,70
13	2,83	0,64
14	2,02	0,45
15	1,85	0,42
16	1,53	0,34
17	1,45	0,33
18	1,27	0,29
19	1,14	0,26
Total varians forklart (%)		99,04

4.1 Korrelasjonen mellom konvensjonell-EBV og PC-EBV

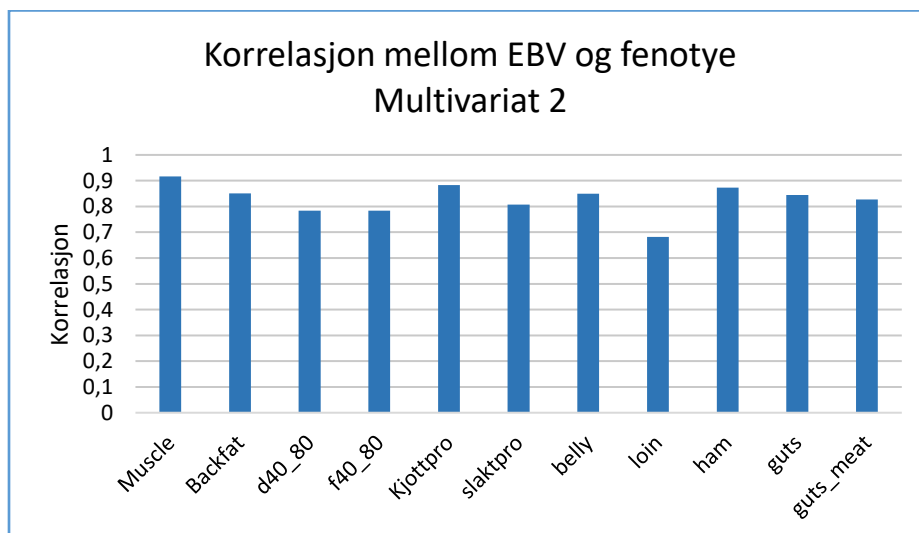
Fra figurene 4, 5, 6 og 7 kan en se korrelasjonen mellom de predikerte avlsverdiene (EBV) ved bruk av den konvensjonelle metoden og PC metoden.

Figur 4 viser korrelasjonen mellom EBV for egenskapene som er i gruppe 1 (tabell 1). Figuren viser at det er forholdsvis høye korrelasjoner mellom de to metodene. For flere av egenskapene er korrelasjonen nesten 90%.



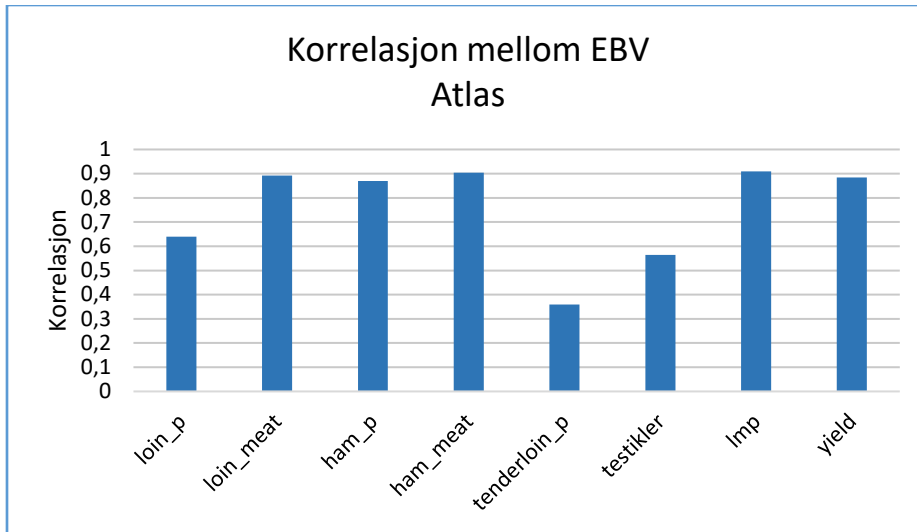
Figur 4: Korrelasjonen mellom avlsverdiene for PC-egenskapene og de konvensjonelle egenskapene. Dette er egenskapene i gruppe 1 (tabell 4).

Korrelasjonen mellom EBV for egenskapene i gruppe 2 er vist i figur 5. Denne viser også mange høye korrelasjoner hvor de fleste er over 80%.



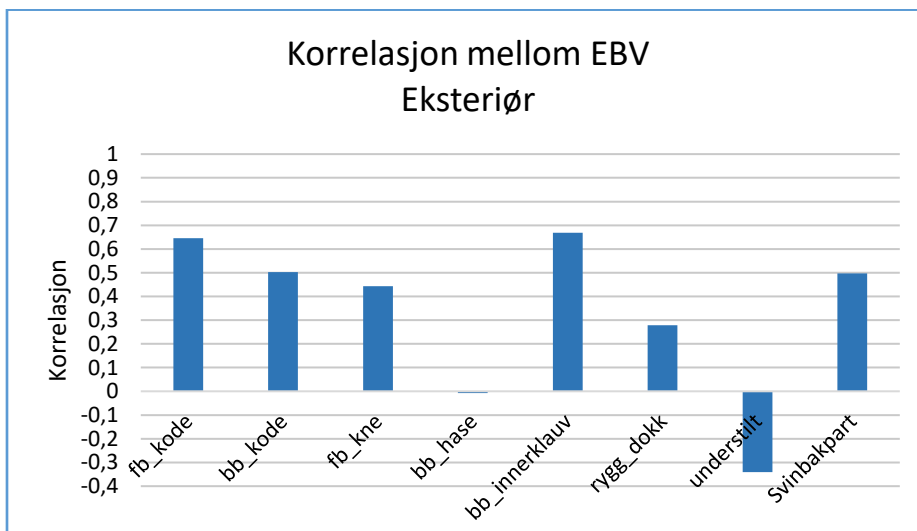
Figur 5: Korrelasjonen mellom avlsverdiene for de konvensjonelle-egenskapene og PC-egenskapene. Dette er egenskapene i gruppe 2 (tabell 4).

Figur 6 viser korrelasjonene mellom EBV for Atlasegenskapene, altså gruppe 3. Det er fem egenskaper som har en korrelasjon på nesten 90%, mens tre egenskaper har lave korrelasjoner.



Figur 6: Korrelasjonen mellom avlsverdiene for de konvensjonelle-egenskapene og PC-egenskapene. Dette er egenskapene i gruppe 3 (tabell 4).

Figur 7 viser korrelasjonene mellom EBV for eksteriøregenskapene, altså i gruppe 4. Denne figuren viser lave korrelasjoner mellom de to metodene. Disse korrelasjonen er vesentlig lavere enn i de tre andre figurene, og for en av egenskapene er den til og med negativ.

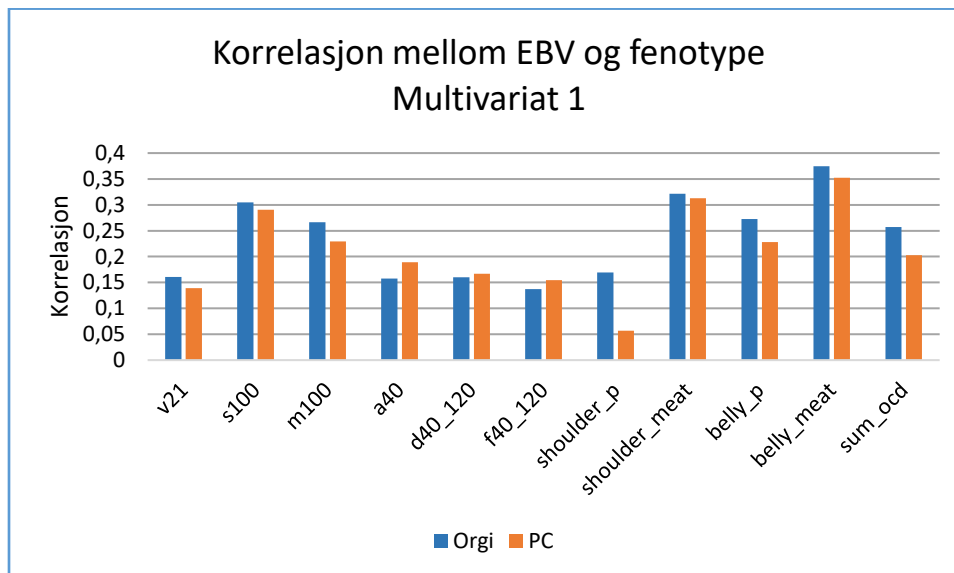


Figur 7: Korrelasjonen mellom avlsverdiene for de konvensjonelle-egenskapene og PC-egenskapene. Dette er egenskapene i gruppe 4 (tabell 4).

4.2 Kryssvalidering

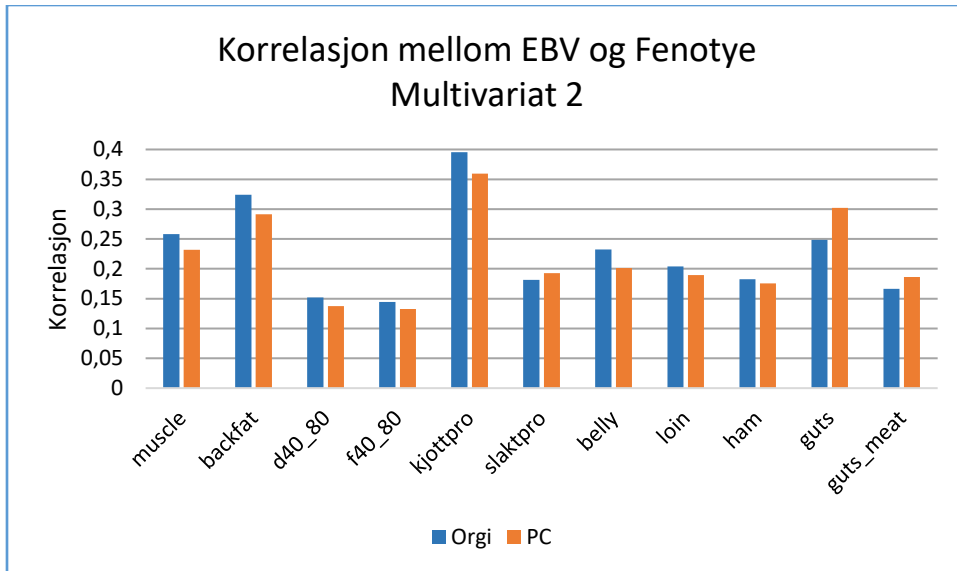
Resultatene fra kryssvalideringen i det første forsøket vises i figur 8, 9, 10 og 11. Disse figurene viser korrelasjonene mellom de predikerte avlsverdiene og fenotypen på de 300 ekskluderte individene.

Fra figur 8 kan en se at korrelasjonene mellom den konvensjonelle-EBV og fenotypene er ganske like korrelasjonene mellom PC-EBV og fenotypene. PC-EBV er stort sett litt lavere enn for konvensjonelle-EBV. Det er tre PC-EBV som er høyere enn konvensjonell-EBV, a40, d40_120 og f40_120. Allikevel er korrelasjonen svært mye lavere for shoulder_p ved bruk av PCA enn uten.



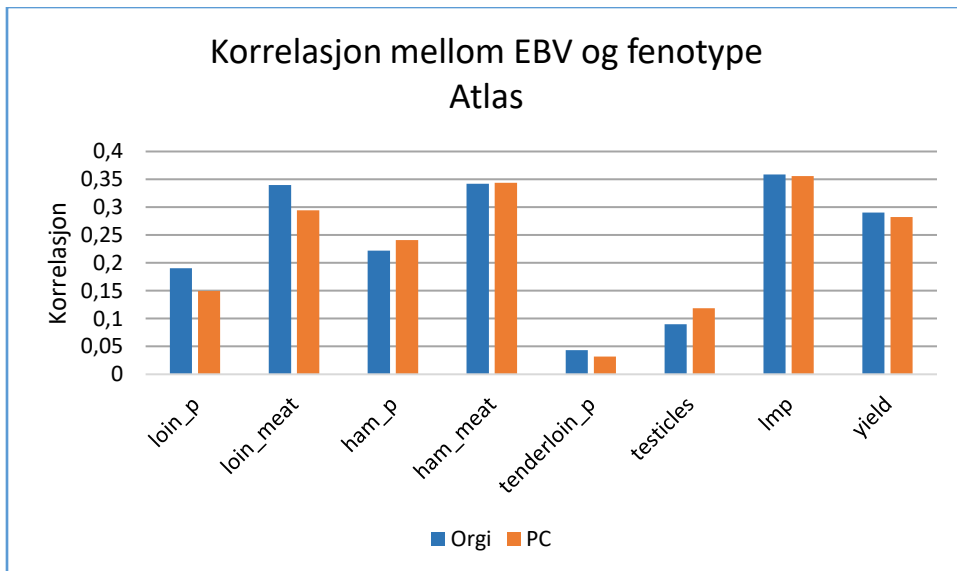
Figur 8: Sammenligning av korrelasjonen mellom avlsverdi og fenotype for den konvensjonelle metoden (orgi) og PC metoden. Dette er egenskapene i gruppe 1 (tabell 1).

Tilsvarende viser figur 9 at det er lite forskjell mellom de to metodene. Som i figur 8 er det noen få egenskaper hvor PC-EBV er litt bedre enn konvensjonell-EBV.



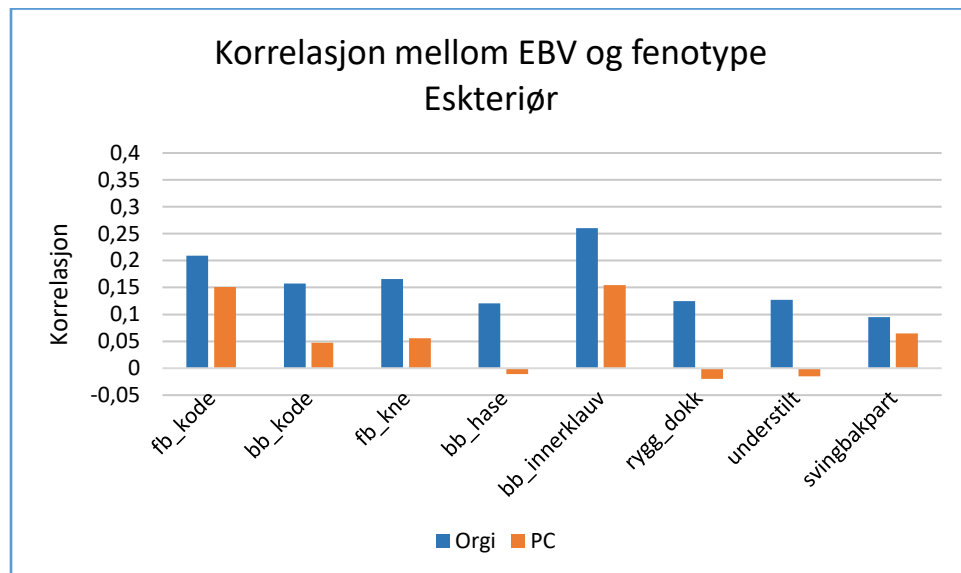
Figur 9: Sammenligning av korrelasjon mellom avlsverdi og fenotype for den konvensjonelle metoden (orgi) og PC metoden. Dette er egenskapen fra gruppe 2 (tabell 4).

Figur 10 viser Atlasegenskapene (gruppe 3). Også her er det svært liten forskjell mellom den konvensjonelle metoden og PC metoden.



Figur 10: Sammenligning av korrelasjonen mellom avlsverdi og fenotype for den konvensjonelle metoden (orgi) og PC metoden. Dette er egenskaper fra gruppe 3 (tabell 4).

Figur 11 viser korrelasjonen mellom EBV og fenotypene i gruppe 4. Denne figuren har ganske annerledes resultat enn de tidligere figurene. Her er alle de konvensjonelle-EBV mye høyere korrelert med fenotype, enn det PC-EBV er. For to egenskaper (fb_kode og bb_innerklauv) korrelerer PC-EBV og fenotype høyt, men allikevel er korrelasjonen mellom konvensjonell-EBV og fenotype høyere. Generelt er korrelasjonene mellom PC-EBV og fenotypene lave i denne figuren.



Figur 11: Sammenligning av korrelasjonen mellom avlsverdi og fenotype for den konvensjonelle metoden (orgi) og PC metoden. Dette er egenskaper fra gruppe 4 (tabell 4).

Tabell 6 gir en oversikt over egenskapene som gjør det bedre ved bruk av PC-EBV.

Tabell 6: PC-EBV som gjør det bedre enn den konvensjonelle-EBV, samt en beskrivelse i prosent hvor mye bedre de gjør det.

Egenskaper	Prosent
A40	3,2
D40_120	0,6
F40_120	1,8
Slaktepro	1,1
Guts	5,4
Guts_meat	2
Ham_p	1,9
Testicles	2,9

4.3 Genetiske korrelasjoner mellom PC

Tabell 7 viser at det er genetiske korrelasjoner mellom PC. Disse verdiene sprer seg fra -0,49 til 0,53. Standardfeilen til korrelasjonsmatrisen er vist i tabell 8.

Tabell 7: Den genetiske korrelasjonen mellom de 19 PC.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1	1																			
2	-0,34	1																		
3	0,23	-0,17	1																	
4	0,13	0,34	-0,43	1																
5	-0,01	0,04	0,16	0,39	1															
6	-0,23	0,12	0,31	0,19	0,41	1														
7	0,03	0,22	-0,06	-0,33	-0,18	-0,06	1													
8	-0,15	-0,03	0,02	0,05	0,00	-0,15	-0,26	1												
9	0,32	-0,09	0,39	-0,17	0,31	0,23	0,07	-0,24	1											
10	0,12	-0,20	0,13	0,10	-0,16	0,00	-0,34	0,10	0,00	1										
11	0,12	-0,06	0,28	0,34	0,18	0,10	-0,31	0,27	-0,42	-0,03	1									
12	0,16	-0,20	-0,02	0,20	-0,20	0,34	-0,49	0,28	0,01	0,06	0,18	1								
13	0,12	0,10	-0,06	0,01	-0,04	0,08	0,30	0,00	-0,26	-0,16	-0,03	0,22	1							
14	-0,02	0,02	0,38	-0,12	0,29	0,26	-0,06	-0,11	0,03	-0,37	0,27	0,11	0,22	1						
15	-0,10	0,02	0,29	-0,37	-0,26	-0,09	0,16	-0,21	-0,06	0,01	-0,24	-0,38	-0,14	0,32	1					
16	0,05	-0,04	-0,36	0,32	0,22	0,49	0,12	-0,31	0,12	-0,17	0,02	0,35	-0,02	0,08	-0,49	1				
17	-0,03	-0,09	0,02	0,27	-0,12	0,53	-0,01	-0,17	0,06	0,04	0,11	0,42	0,22	-0,13	-0,09	0,20	1			
18	-0,24	0,07	-0,16	0,00	0,01	0,08	-0,28	0,08	-0,09	-0,32	-0,34	0,17	0,09	-0,11	0,23	-0,26	0,09	1		
19	0,01	0,34	0,21	-0,32	-0,33	-0,25	0,42	-0,06	0,19	0,30	-0,24	-0,33	0,17	-0,39	-0,08	-0,39	-0,05	-0,33	1	

Tabell 8: Standardfeilen til den genetiske korrelasjonen mellom de 19 PC.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0																		
2	0,17	0																	
3	0,21	0,18	0																
4	0,22	0,18	0,22	0,00															
5	0,25	0,21	0,26	0,27	0														
6	0,20	0,17	0,21	0,21	0,24	0													
7	0,19	0,16	0,19	0,19	0,22	0,18	0												
8	0,19	0,17	0,21	0,21	0,24	0,19	0,18	0											
9	0,17	0,15	0,17	0,19	0,22	0,17	0,16	0,16	0										
10	0,22	0,18	0,23	0,23	0,27	0,22	0,20	0,21	0,19	0									
11	0,22	0,19	0,22	0,23	0,27	0,23	0,20	0,21	0,18	0,24	0								
12	0,20	0,18	0,22	0,23	0,25	0,20	0,18	0,20	0,19	0,22	0,23	0							
13	0,28	0,24	0,29	0,31	0,34	0,28	0,25	0,27	0,25	0,30	0,31	0,28	0						
14	0,26	0,23	0,27	0,29	0,30	0,26	0,23	0,26	0,24	0,26	0,28	0,26	0,35	0					
15	0,17	0,15	0,18	0,18	0,22	0,17	0,16	0,16	0,15	0,19	0,19	0,16	0,24	0,21	0				
16	0,21	0,18	0,21	0,22	0,26	0,21	0,19	0,20	0,19	0,23	0,23	0,20	0,30	0,27	0,17	0			
17	0,20	0,17	0,21	0,22	0,25	0,19	0,18	0,19	0,18	0,22	0,23	0,22	0,30	0,26	0,18	0,21	0		
18	0,17	0,14	0,18	0,18	0,21	0,17	0,15	0,16	0,15	0,18	0,18	0,17	0,23	0,22	0,14	0,17	0,17	0	
19	0,22	0,18	0,23	0,23	0,26	0,20	0,19	0,21	0,19	0,23	0,24	0,21	0,29	0,30	0,18	0,23	0,22	0,18	0

5 Diskusjon

Da datamaterialet ble laget for denne masteroppgaven hadde ikke Norsvin mer informasjon om Atlasegenskapene enn fra 1830 dyr. Et større datasett ville vært gunstig, for dette vil bidra til mer informasjon og kanskje gi et bedre resultat. Antall dyr som har fenotype på atlagegenskapen øker, så i fremtiden vil dette bli mulig.

Det ble ikke tatt med noen flere tilfeldige effekter enn individ, for å redusere antallet varianskomponenter som trengtes å estimeres. Tilfeldige effekter er i midlertidig viktige å ta hensyn til fordi om disse fjernes kan det føre til en overestimering av den additive genetiske variansen. Dette kan medføre at det blir større variasjon mellom avlsverdiene, som kan gi inntrykk av at enkelte avlsverdier er bedre enn de faktisk er. Derfor bør en inkludere de tilfeldige effekter ved neste utprøving av PC.

Tabell 5 viser at en valgte 19 PC, for disse hadde en egenverdi større enn 1,0. Det ble beregnet varianser og avlsverdier på de 19 PC i DMU. Tidligere ble det sagt at det ikke var mulig å beregne varianskomponenter for mer enn 11 egenskaper i DMU, ved bruk av den konvensjonelle metoden, mens det gikk for hele 19 PC. Dette kan tyde på at det er bra kondisjon på matrisen med de 19 PC. Dermed har PCA gjort det mulig å estimere og predikere avlsverdier på flere PC enn det en får til på de originale egenskapene.

Korrelasjonen mellom konvensjonell-EBV og PC-EBV er generelt høye i både figur 4 og 5, men er litt mer varierende i figur 6 og 7. I figur 6 har fem av EBVene en korrelasjon på nesten 90%, mens de resterende tre er ganske lave. For figur 7 er det ingen egenskaper som er korrelert mer enn 70%, og i denne gruppen er det til og med noen negativt korrelert EBV.

Kryssvalideringen gjort for de tre første gruppene, figur 8, 9 og 10, viser tydelig at de to ulike metodene beregner veldig tilsvarende EBVer. Faktisk er noen av PC-EBVene høyere korrelert med fenotypen enn konvensjonell-EBVene. Grunnen til at noen PC-EBVene er høyere korrelert kommer antageligvis av at korrelerte egenskaper styrker oppunder hverandre, Lynch and Walsh (1998) påpekte at ved bruk av informasjon fra korrelerte egenskaper blir det mer riktige estimater. I tillegg brukes informasjonen fra dataen til å beregne færre varianser via PCA, enn ved den konvensjonelle metoden. Dette gjør at informasjonen ikke blir fordelt på et like stort antall egenskaper og sånn gi sikrere estimater. Figur 11 kommer derimot ikke så bra ut for PC-

EBV, dette kan antagelig ha med at eksteriøregenskapene har liten sammenheng med produksjonsegenskapene og at det er lite av informasjonen om eksteriøregenskapene som har kommet med inn i de 19 PC.

Tabell 6 gir en liten oversikt over hvilke egenskaper som gjør det bedre ved bruk av PC-EBV. De tre egenskapene som gjør det best her er; alder ved 40kg (a40), antall kg som er mage (guts) og testikler. Egenskapen a40 er med i avlsmålet, og anses derfor å være en av de viktigere egenskapene (1% i total avlsverdi). Denne tabellen viser at det ikke er store forskjellen, men allikevel verdt å merke seg at tre av egenskapene som er på lista er med i dagens avlsmål, i tillegg til egenskapene alder ved 40kg er det egenskapene dager brukt for å vokse fra 40 til 120 kg (d40_120) (4% i total avlsverdi) og mengden fôr brukt for å vokse fra 40 til 120 kg (f40_120) (7% i total avlsverdi).

Tabell 7 viser at det er genetiske korrelasjoner mellom alle de 19 PC. Dette betyr at ved bruk av PCA vil en utnytte informasjon om korrelert egenskaper. PCA inkluderer informasjon om korrelasjonen mellom egenskapene som vil kunne bidra mer, fordi uten PCA settes korrelasjonen til 0. Tabell 8 viser standardfeilen for matrisen til de genetiske korrelasjonene.

Som nevnt har ikke eksteriøregenskapene kommet godt ut i beregningene. Altså ser det ut til at PCA henter ut lite informasjon fra disse egenskapene. Dette har med at eksteriøregenskapene ikke har noen vesentlige sammenhenger med de andre egenskapene valgt i denne oppgaven. Men en skal ikke se bort ifra at eksteriøregenskapene har en vesentlig effekt på egenskaper som holdbarhet og hold etter avvenning. Så selv om eksteriøregenskapene var overflødige her, kan de være vesentlig å ha med når det er andre egenskaper enn produksjonsegenskaper som studeres.

I denne oppgaven valgte en å bruke PC som hadde en egenverdi høyere enn 1,0. Fra resultatene kan det tyde på at dette var fornuftig, side de 19 PC forklarte det meste av variasjonen i dataen. Det ble også forsøkt å gjøre det samme med bare 10 PC, men da viste det seg at det ble kuttet for hardt, og for lite informasjon ble tatt med. Dette ga generelt dårlige resultater i kryssvalideringen, med unntak av noen egenskaper som faktisk gjorde det tilsvarende som den konvensjonelle metoden. Det var egenskapene muskeltykkelse i mm målt ved siste ribbein (m100), fôr brukt for å vokse fra 40 til 120 kg (f40_120), mengden muskler på tidspunktet grisen ble veid (muscle) og hvor mange kg testikler grisen har (testicles). Her kan det tyde på at det er fornuftig å inkludere PC med en egenverdi høyere enn 1,0.

Kryssvalideringstesten ble bare gjort på 296 individer, noe som er et veldig begrenset antall. Et lite datasett kan redusere nøyaktigheten til de predikerte avlsverdiene (Kennedy & Trus, 1993), derfor vil et større antall vil gi en sikrere sammenligning. Derfor burde forskjellen mellom metodene videre testes på et større datasett.

I denne oppgaven har PC blitt brukt for å estimere varianskomponenter og avlsverdier. Gjennom denne metoden har en fått avlsverdiprediksjoner som i noen tilfeller er bedre enn den konvensjonelle metoden, antagelig på grunn av kovarianser mellom egenskapene. Derfor er det naturlig at PCA også kan benyttes for å beregne varianskomponenter. Det vil si at en gjør en PCA på datasettet som tidligere gjort, men bare gjør en variansestimering i DMU. Deretter kan variansene bli tilbakekonvertert, for så at en får varianskomponenter på alle de originale egenskapene tatt med inn i PCA. Dette vil bety at det vil være verdier i kovariansmatrisen istedenfor at det står 0. Sammenhengene mellom egenskapene vil bli tatt hensyn til, noe som ikke skjer i dag siden de estimerer variansene på en gruppe av gangen. Disse varianskomponentene kan så brukes når en skal predikere avlsverdier via den konvensjonelle metoden.

Norsvin benytter seg av økonomiske vekter for de egenskapene som er med i avlsmålet. Hvis det blir gjennomført PCA på egenskaper som er med i avlsmålet, kan de økonomiske vektene bli brukt som normalt.

I praksis benytter Norsvin seg av geonomisk slektskap (Andersen-Ranberg, 2018), og har som regel denne informasjonen med inn i DMU. Det har ikke blitt gjort i denne oppgaven, da dette er med på å gjøre utregningene i DMU enda tyngre.

6 Konklusjon

Oppgaven viser at det via PCA er mulig å beregne avlsverdier for hele 41 egenskaper om gangen. PCA-tilnærmingen viser ingen klar sammenheng mellom eksteriøregenskapene og produksjonsegenskapene. Kryssvalideringen viser at noen egenskaper predikeres bedre via PCA enn ved den konvensjonelle tilnærmingen. Det blir bedre prediksjoner for noen av egenskapene siden informasjonen brukes til å bestemme færre varianskomponenter via PCA, enn ved den konvensjonelle metoden. I tillegg blir det kovarianser på tvers av alle egenskapene, hvor dette ikke er tilfelle for gruppekjøringene. Det ser ut til at det ekstra informasjon å hente ved å bruke PCA, derfor burde denne metoden videre testes.

Kilder

- Anderen-Ranberg, I. M. (2016). *Avlsmål for Norsvin Landsvin*: Norsvin (16.03).
- Andersen-Ranberg, I. M. (2016). *Forelesning i Avlsarbeidet På Gris ved IHA* Ås: Norges miljø- og biovitenskapelige universitet (11.01).
- Andersen-Ranberg, I. M. (2018). *Personlig samtale; Avlsverdiberegning* (30.04).
- Andonov, S., Ødegård, J., Svendsen, M., Ådnøy, T., Vegara, M. & Klemetsdal, G. (2013). Comparison of random regression and repeatability models to predict breeding values from test-day records of Norwegian goats. *Dairy Science*: 1834-1843.
- Brekke, M. & Kolbenstvedt, A. (2014). CT. Tilgjengelig fra: <https://sml.snl.no/CT>.
- Bruksbesetninger*. (2013). Norsvin. Tilgjengelig fra: <https://norsvin.no/Avl/Avlsprogram/Avlssystemet/Bruksbesetninger> (lest 14.03).
- Dagnachew, B. S., Meuwissen, T. H. E. & Ådnøy, T. (2012). Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats. *American Dairy Science Association* (96): 5933-5942.
- Dallas, G. (2013). *Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction*. UK based Information Engineer/Internet Social Scientis: George Dallas. Tilgjengelig fra: <https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/> (lest 14.03).
- Esbensen, K. H. (2000). *Multivariate Data Analysis -in practice*. 4 utg.: CAMO ASA. 600 s.
- Foredlingsbesetninger*. (2012). Norsvin. Tilgjengelig fra: <https://norsvin.no/Avl/Avlsprogram/Avlssystemet/Foredlingsbesetninger> (lest 05.03).
- Formeringsbesetninger*. (2013). Norsvin. Tilgjengelig fra: <https://norsvin.no/Avl/Avlsprogram/Avlssystemet/Formeringsbesetninger> (lest 14.03).
- Gangsei, L. E. & Kongsro, J. (2015). Automatic segmentation of Computed Tomography (CT) images of domestic pig skeleton using a 3D expansion of Dijkstra's algorithm. *Computers and Electronics in Agriculture*: 191-194.
- Gjerlaug-Enger, E. J. B. (2011). *Genetic analyses of meat, fat and carcass quality traits measured by rapid methods* Ås: Department of Animal and Aquacultural Sciences,

Norwegian University of Life Sciences, Genetiske analyser av kjøtt-, fett- og slaktekvalitetsegenskaper målt med hurtigmatoder. 171 s.

Ingris. (2016). Animalia. Tilgjengelig fra:

<https://www.animalia.no/no/Dyr/husdyrkontrollene/ingris/> (lest 14.03).

Kennedy, B. W. & Trus, D. (1993). Considerations on genetic connectedness between management units under an animal model. *Journal of Animal Science*, 71 (9): 2341-2352.

Kongsro, J. (2015). Under The Surface Of A Pig. Tilgjengelig fra:

[https://www.norsvin.no/Avl/Avlsprogram/Raanetest/Norsvins-bruk-av-datatomografi-CT-scanning/\(language\)/nor-NO](https://www.norsvin.no/Avl/Avlsprogram/Raanetest/Norsvins-bruk-av-datatomografi-CT-scanning/(language)/nor-NO)

Kongsro, J. (2017). *Kartlegging av grisens kropp ved bruk av atlas*. Blogg: Jørgen Kongro.

Tilgjengelig fra: <https://jorgenkongsro.wordpress.com/2017/10/18/kartlegging-av-grisens-kropp-ved-bruk-av-atlas/> (lest 15.03).

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, Mass: Sinauer Associates.

Madsen, P. & Jensen, J. (2013). A User's Guide to DMU. 32. Tilgjengelig fra:

http://webcache.googleusercontent.com/search?q=cache:Q_gnF5v2WDYJ:dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf+&cd=1&hl=no&ct=clnk&gl=no.

Martinsen, K. H., Meuwissen, T. H. E., Ødegård, J. & Olsen, D. (2016). *Genetic analyses of feed efficiency traits in pigs*. Philosophiae doctor (PhD) thesis. Ås: Norwegian University of Life Sciences Department of Animal and aquacultural Sciences 100 s.

MathWorks. (2017). *MATLAB*. 9.3.0.713579 utg., 2018, 24.04.

NCSS. (2018). *Principal Components Analysis*. Tilgjengelig fra: [https://ncss-wpengine.netdna-](https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Principal_Components_Analysis.pdf)

[ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Principal_Components_Analysis.pdf](https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Principal_Components_Analysis.pdf) (lest 03.05).

SAS. (2013). User's Guide 9.3.

Vangen, O., Steine, T., Olesen, I. & Hårdnes, T. (1994). *Avlslære*. Oslo: Landbruksforlaget



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

Postboks 5003
NO-1432 Ås
Norway