

"Teknikken" i testdagsmodellen

Jørn Pedersen, Afdeling for Specialviden, Dansk Kvæg

Nordisk avlsværdiurdering – kort opsummering

3 racegrupper:

- Nordisk Holstein: Dansk, Svensk og Finsk Holstein samt Dansk Rød Holstein
- Røde racer: RDM, SRB, Finsk Ayrshire og Finn Cattle
- Jersey: Dansk Jersey

Datagrundlag:

- Sverige: 1.-3. laktations 305d fra 1990 og frem
- Danmark: 1.-3. laktations TD – ydelser fra 1990 og frem
- Finland:
 - 1.-9. laktations TD-ydelser fra 1988 og frem
 - Både Holstein og Ayrshire data indgår i både den "røde" evaluering og i Holstein evalueringen

Model

- Multi-trait for egenskaberne: Mælk, protein og fedt
- Multi-trait for 1., 2. og 3. laktation. Finske 3.-9. laktationer evalueres med gentagelsesmodel
- TD-modellerne er næsten ens for Danmark og Finland
 - Besætningseffekt er "Random" i Finske data
 - Krydsningseffekter er selvfølgelig forskellige for racerne
- For 305d anvendes den nationale Svenske model

Testdagsmodellen

Den nordiske Testdagsmodel er:

- "Multi-laktation" model, hvor hver laktation bliver anset for at være een egenskab.
- "Multi-trait" model, hvor egenskaberne mælk, fedt og protein bliver analyseres samtidigt
- Den tilfældige effekt af dyr, avlsværdierne, antages at være den samme i Sverige, Finland og Danmark. Derimod bliver der tilladt forskellig tilfældig miljømæssig variation i de 3 lande.

I en testdagsmodel indgår systematiske miljøeffekter og tilfældige effekter, som både er miljømæssige og genetiske effekter. Til forskel fra de systematiske effekter kræver de tilfældige effekter kendskab til størrelse af varianser og kovarianser

Ud over den klassiske modelformulering med systematiske og tilfældige effekter omfatter beregningen korrektion for heterogen varians.

Systematiske miljøeffekter

I den danske og finske testdagsmodel indgår følgende systematiske miljøeffekter (den svenske model for 305-dages ydelser ligner den model, som blev anvendt i Danmark indtil april 2006)

- Besætning x kontroldag (tilfældig i Finland – derfor indgår også kontrolår x måned)
- "Laktationskurve" pr. kontrolmåned x 4-årsperiode x 3 aldersgrupper i 1. laktation
- "Laktationskurve-korrektion" pr besætning x 4-årsperiode
- Kælvingensalder pr 5-årsperiode
- Drægtighedsdage
- Goldperiodens længde i foregående laktation, dvs. kun 2. og 3. laktation
- Heterosis og phantom-forældregrupper

Besætning x kontroldag

En af de mest markante fordele ved en testdagsmodel frem for en 305-dages model er at det bliver muligt at foretage en meget præcis korrektion for begivenheder, der påvirker ydelserne i en besætning på en bestemt testdag (foderskift, personaleskift, vejrskift mv.). I den nordiske testdagsmodel beregnes effekten af testdag(kontroldag)

- For 1. laktation separat
- For 2. og 3. laktation samlet

Laktationskurver og laktationskurve pr besætning

For at ydelser kan sammenlignes inden for samme kontrol dag er det nødvendigt at der bliver taget højde for at køerne er på forskellige laktationsstadier. Derfor bliver der taget højde for laktationskurvens form. Den kan være forskellig afhængigt af årstid, år og for 1. kalvskøer også kælvningsalderen

- 1. laktation: 12 måneder * tidsperioder (nu 5 perioder) * 3 aldersgrupper, dvs. for hver egenskab er der $12 * 5 * 3 = 180$ laktationskurver (niveauet kan være forskelligt fra år til år inden for tidsperiode)
- 2. og 3. laktation: 12 måneder * tidsperioder (nu 5 perioder), dvs. for hver egenskab og laktation er der 60 forskellige laktationskurver (niveauet kan være forskelligt fra år til år inden for tidsperiode)

Laktationskurverne er hver især bestemt af en 5-parameter funktion:

$$\text{dagsydelse} = a + b * \text{dag} + c * \text{dag}^2 + d * e^{k_1 * \text{dag}} + f * e^{k_2 * \text{dag}}$$

hvor

- dag er afstand fra kælving (1 – 365 dage) konverteret til et tal mellem -1 og +1
- a, b, c, d, f er de parametre, der bestemmer kurvernes form.
- k_1 og k_2 er konstanter (fastsat med dataanalyser)
- k_1 er ens for alle racer og alle laktationer og er fastsat til -0,05 for mælk og -0,04 for protein og fedt. k_1 bliver også anvendt i de funktioner der beskriver udvikling af avlsværdier og miljøeffekter hen over laktationen
- For k_2 er der lidt flere forskelle mellem racer, egenskaber og laktationer (Tabel 1)

De 3 første led i funktioner er en "klassisk" andengradsfunktion, der vil beskrive laktationskurven som en parabel. De 2 sidste led modificerer kurvens forløb i den første del af laktationen. Efter dag 50 – 100 i laktationen vil de være uden betydning

Disse laktationskurver tager ikke hensyn til at laktationskurvens form kan være forskellig fra besætning til besætning. Derfor indgår der en besætningsafhængig modifikation af laktationskurve.

For hver besætning beregnes en korrektionsværdi
 korrektion af dagsydelse = $b \cdot \text{dag}$

Hvis b er positiv er det fordi besætningen generelt har fladere laktationer end gennemsnittet, dagsydelseerne er forholdsvis høje i slutningen af laktationen – og omvendt, hvis b er negativ

Tabel 1. k_2 -konstant i funktionen til beskrivelse af fænotypiske laktationskurve. k_1 er ens for alle racer og alle laktationer og er fastsat til -0,05 for mælk og -0,04 for protein og fedt

Egenskab	Laktation	Røde racer	Holstein	Jersey
Mælk	1	-0,01	-0,01	-0,01
	2	-0,01	-0,01	-0,02
	3	-0,01	-0,01	-0,02
Protein	1	-0,09	-0,09	-0,09
	2	-0,09	-0,09	-0,05
	3	-0,09	-0,09	-0,05
Fedt	1	-0,02	-0,02	-0,02
	2	-0,02	-0,02	-0,08
	3	-0,02	-0,02	-0,08

Kælvningsalder

Der er velkendt, at alder ved kælvning har indflydelse på ydelsesniveauet i både 1. 2. og 3. laktation. I hver laktation bliver der taget hensyn til månedsvise aldersforskelle

I den røde racegrupper viste det sig, at aldersudviklingen var forskellig for de forskellige delpopulationer. Aldersudviklingen var stærkere hos køer med mange ABK- eller HOL-gener end hos køer med f.eks. mange SRB- eller FAY-gener. Forskellene var så store, at det var nødvendigt at tage højde for disse forskelle i testdagsmodellen.

Drægtighedsdage

I 305-dages modeller til beregning af avlsværdital indgår normalt en korrektion for effekt af samtidigt kælvingsinterval (eller tomperiode) for at tage højde for drægtighedens negative virkning på ydelsen. Testdagsmodellen giver mulighed for en langt mere avanceret korrektion, nemlig effekt af drægtighedsdage (hvor lang tid koen har været drægtig) på dagsydelse.

- 0 – 119 dage: Det viste sig, at hos køer der havde være drægtige mindre end 120 dage har drægtigheden ingen virkning på ydelsen. Derfor ikke-drægtige køer (0 dage drægtige) og køer der er 1-120 dage henne i drægtigheden slået sammen i en gruppe
- 120 – 239 dage: Effekten bliver beregnet for 10-dages perioder
- 240 og mere: Samlet i en klasse (de fleste køer er goldet på det tidspunkt)

Golddage

I testdagsmodellen indgår effekt af antal golddage i foregående laktation (kun 2. og 3. laktation), fordi goldningstidspunktet har indflydelse på ydelsen i den efterfølgende laktation. For goldperioder under 56 bliver effekten bliver bestemt for 7-dages perioder. For goldperioder 56 dage og derover beregnes en samlet effekt.

Antal golddage bliver fastsat på 2 måder

- Indberettet golddag (kun i Danmark)
- Midtpunkt mellem sidste kontrol med ydelse og efterfølgende kontrol/kælvning/afgang

Effekten bliver bestemt særskilt for de 2 grupper af registreringer

Heterosis og rekombination

Der indgår effekt af heterosis for alle 3 racegruppe. For Holstein og Jersey indgår effekten af heterosis på samme måde som i den model, vi anvendte indtil april 2006. Effekt af rekombination er ikke inkluderet.

I den røde racegruppe viste det sig også at være nødvendigt at inddrage effekten af rekombination. Recombination er en effekt der medvirker til at nedtone heterosiseffekten ved krydsning over flere generationer. Rent genetisk bliver den negative effekt forklaret ved at positive gen-kombinationer bliver brudt op ved længere tids krydsning.

Effekten af heterosis bliver beregnet ud fra graden af heterozygoti, som er:

$$\text{Race1 i far} * \text{race2 i mor} + \text{race2 i far} * \text{race1 i mor}$$

Effekten af rekombination bliver beregnet ud fra rekombinationskoefficienten, som er:

$$\text{Race1 i far} * \text{race2 i far} + \text{race1 i mor} * \text{race2 i mor}$$

I den røde racegruppen er der mange forskellige "racer" repræsenteret. Derfor var det nødvendigt at modificere metoden.

- Derfor bestemmes heterosis og rekombination som en generel effekt inden for hvert land.
- De forskellige kombinationer bliver derefter bestemt som en afvigelse fra den generelle heterosis- eller rekombinationseffekt, som en tilfældig effekt (effekterne "regresses" mod det generelle gennemsnit)

Holstein: Heterosiseffekter

- HF x opr. SDM
- HF x opr. DRH
- HOL x FAY: kun i finske data
- Intern FAY-heterosis: Kun FAY-køer i finske data

Jersey: Heterosiseffekter

- DJ x USJ
- DJ x NZJ

RDM

Generel effekt af heterosis og rekombination

Danmark: Tilfældige heterosiseffekter og rekombinationseffekter

- RDM x ABK
- RDM x HOL
- RDM x (SRB + FAY + NRF)
- ABK x HOL
- ABK x (SRB + FAY + NRF)

Sverige: Tilfældige heterosiseffekter og rekombinationseffekter

- RDM x SRB
- ABK x SRB
- SRB x CAN
- SRB x FAY
- SRB x NRF
- FAY x NRF

Finland: Tilfældige heterosiseffekter.

SRB x NRF og FAY x NRF heterosis indgår ikke i det finske afsnit af modellen

- SRB x FAY
- FAY x CAN
- FAY x HOL

Genetiske grupper (*Phantom-parent*)

Modellen tage højde for generens oprindelse med hensyn land og årgang.

- Der er en vigtig effekt fordi heterosis indgår i modellen. Når heterosis indgår, skal de involverende racer og også være en del af modellen. Genetiske grupper er i princippet racegrupper blot med en ekstra findeling med hensyn til tid.
- Dernæst er genetiske grupper også vigtige, når der har været import af arvemateriale over en lang årrække fra populationer med stor genetiske fremgang. De genetiske grupper opfanger avlsfremgangen i eksportlandet, samt forskellen mellem den hjemlige population og det selekterende importerede avlsmateriale.

Et af delresultaterne under udviklingen af den nordiske model var at resultaterne for de genetiske grupper blev meget stabil, når de blev defineret som "tilfældige effekter", dvs. at de bliver regresseret mod et fælles gennemsnit

Korrektion for heterogen varians (HV-korrektion)

Det er vigtigt at korrigerer for heterogen varians. For avlsværdiurderingen generelt er det mest vigtig at tage højde den heterogene varians, som opstår ved at variati-

onen er øget med tiden. Specielt ved udvælgelsen af avlskøer spiller det også en rolle, at der bliver taget højde for, at variationen inden for besætninger kan være meget forskellig

I den model som blev anvendt indtil april 2006, blev data prækorrigeret for heterogen varians, men specielt korrektionen for forskelle mellem besætninger var ret svag. I den nye model er denne korrektion stærkere, men til gengæld er det en "simultan" korrektion, således at de forskelle i variation, der skyldes forskelle i avlsmæssig variation ikke bliver fjernet.

Med "simultant" forstås her en iterativ proces:

- Først beregnes avlsværdier uden HV-korrektion
- På det grundlag foretages den første HV-korrektion
- Derefter genberegnes avlsværdier
- Derefter bliver der foretaget en ny HV-korrektion
- Processen forsættes indtil HV-korrektionerne forbliver uændret fra runde til runde

Til HV-korrektionen hører en række forudsætninger og parametre. Det drejer sig om:

- Basisår skal defineres. Basisåret er i alle tre racegrupper køer født i 1.1.2001-30.6.2003. Det indebærer, at variansen i alle år blive justeret så den er lig med basisårets variation.
- "Rho"-faktorer, som beskriver sammenhængen mellem en testdag og den næste testdag, dvs. i hvor høj grad vi forventer at variansen ved de to testdage er ens.
- Basis-varianser. Disse værdier styrer og fastholder det indbyrdes forhold mellem tilfældig miljøvariation for egenskaber og for lande.

Effekt af dyr / Avlsværdier (tilfældig effekt)

For beskrive udviklingen i avlsværdier hen over laktationen er der for hver egenskab og laktation anvendt en 4-parameterfunktion, som i meget høj grad ligner den, som blev anvendt for den fænotypiske laktationskurve. Det skal fremhæves, at der ikke er tale om en laktationskurve, men noget det kunne kaldes en "avlsværdikurve".

Der er (endnu) ikke påvist nogen logisk årsag til at de 2 funktioner skal være identiske, men praksis har vist det regneteknisk har en gunstig virkning (hurtigere konvergens ved løsning)

$$\text{EBVdagsydelse} = a + b * \text{dag} + c * \text{dag}^2 + d * e^{k_1 * \text{dag}}$$

hvor

- dag er afstand fra kælvning (1 – 365 dage) konverteret til et tal mellem -1 og +1
- a, b, c, d er de parametre der bestemmer kurvernes form.
- k_1 er den samme konstant som blev anvendt for den fænotypiske (fastsat med dataanalyser – tabeller mangler). k_1 er ens for alle racer og alle laktationer og er fastsat til -0,05 for mælk og -0,04 for protein og fedt

For hvert dyr giver denne model 4 avlsværdier (a, b, c og d) pr. laktation og pr. egenskab, i alt 36 avlsværdier pr. dyr. Mange af disse avlsværdier vil være meget højt korrelerede.

Både på grund af de mange avlsværdier og de høje korrelationer vil det regneteknisk blive en meget langsommelig proces at anvende denne funktion direkte.

Antallet af avlsværdier der skal beregnes er derfor reduceret med en metode, der tager udgangspunkt i den statistiske teori om "principale komponenter". Ved at reducere antallet af avlsværdier fra 36 til i alt 14 bibeholdes stort set al information. Efter reduktionen er antallet af avlsværdier:

- 6 samlet for mælk, protein og fedt i 1. laktation
- 8 samlet for mælk, protein og fedt i 2. og 3. laktation

Efter beregning af de 14 avlsværdier bliver de konverteret tilbage til 36 avlsværdier. Når vi kender de 36 avlsværdier kan vi derefter bruge formlen til beregning af avlsværdier, først for dagsydelse, og derefter for f.eks. 305 dagesydelse.

Genetiske parametre

For de tilfældige effekter forudsættes det altid, at vi kender variationen. For de genetiske effekter drejer de sig om genetiske varianser og covarianser (heritabiliteter og korrelationer). I dette tilfælde drejer det sig om de genetiske parametre for de 14 avlsværdier i modellen, men præsentation af disse parametre eller de genetiske parametre for de 36 avlsværdier vil give praktisk mening. Derfor præsenteres kun de genetiske parametre for 305d ydelser (tabel 2, 3 og 4)

Tilfældigt miljø

I testdagsmodellen er den tilfældige, uforklarlige miljøeffekt opdelt i to:

- Permanent miljø inden for laktation. Der er den miljøeffekt som er fælles for hele laktationen (For finske data med 3-9 laktation i en gentagelsesmodel indgå 2 typer permanent miljø)
- Tilfældigt miljø på den enkelte kontolddag

Permanent miljø inden for laktation

Denne effekt bliver ligesom avlsværdierne beskrevet af en 4-parameterfunktion. Der er tale om en helt tilsvarende funktion

$$\text{Permanent miljø for dagsydelse} = a + b * \text{dag} + c * \text{dag}^2 + d * e^{k_1 * \text{dag}}$$

hvor

- dag er afstand fra kælvning (1 – 365 dage) konverteret til et tal mellem -1 og +1
- a, b, c, d er de parametre der bestemmer kurvernes form.
- k_1 er den samme konstant, som blev anvendt for den fænotypiske (fastsat med dataanalyser – tabeller mangler). k_1 er ens for alle racer og alle laktationer og er fastsat til -0,05 for mælk og -0,04 for protein og fedt

For hvert dyr giver denne model 4 resultater (a, b, c og d) pr. laktation og pr. egenskab, i alt 36 effekter pr. dyr. Mange af disse effekter vil være meget højt korrelerede. Både på grund af de mange effekter og de høje korrelationer vil det regneteknisk blive en meget langsommelig proces at anvende denne funktion direkte.

Antallet af effekter, der skal beregnes er derfor reduceret på samme måde som det var tilfældet med avlsværdierne. Ved at reducere antallet af effekter fra 36 til i alt 18 bibeholdes stort set al information. Efter reduktionen er antallet af effekter:

- 6 samlet for mælk, protein og fedt i 1. laktation
- 6 samlet for mælk, protein og fedt i 2. laktation
- 6 samlet for mælk, protein og fedt i 3. laktation

Normalt vil vi ikke være interesseret i, at beregne den permanent miljøeffekt for dagsydelse, men det kunne lade sig gøre efter samme skabelon som ved beregning af avlsværdital

Tilfældigt miljø på den enkelte kontroldag

Tilfældigt miljø på en enkelt kontroldag forudsættes at være konstant over hele laktationen.

Miljømæssige parametre

For de tilfældige effekter forudsættes det altid, at vi kender varianserne og covarianserne. Som for de genetiske parametre gælder det, at præsentation af værdierne for de 18 reducerede parametre eller de 36 oprindelige parametre vil være meningsløs. De miljømæssige parametre er forskellige fra land til land

Udenlandsk information

Udenlandsk avlsværdital bliver indregnet i avlsværditalene

- De udenlandske avlsværdital bliver omregnet til ydelsesafvigelse ("deregressed")
- Disse ydelsesafvigelser indgår i beregningen på lige fod med andre ydelsesdata (på samme måde som svenske 305-dages ydelser)
- Informationer tilknyttes "besætningsklasser" afhængigt af hvilke land de udenlandske informationer stammer fra. Derudover fortages der selvfølgelig ikke korrektion af data.
- Sikkerhed på de udenlandske avlsværdital bestemmer, hvilken vægt der bliver lagt på de udenlandske informationer."
- Der bliver ikke foretaget heterogen varianskorrektion af de udenlandske informationer. I praksis betyder det at, avlsværdiberegninger skal køres 2 gange
 - Første en gang uden udenlandsk avlsværdital. I denne beregning fastlægges størrelsen af HV-korrektionen.
 - Anden gang med udenlandsk information inkluderet, men uden HV-korrektion. I stedet bliver data prækorrigeret for heterogen varians med de faktorer som blev beregnet i første omgang

Avlsværdiurderingen trin for trin

- Danske data trækkes ud af databasen, og der foretages en dansk editering af data.
- Data afleveres til NAV separat for RDM, for Jersey og for SDM-DH og DRH
- Hos NAV samles data fra de 3 lande og kombineres med den nordiske afstammingsfil
- Det bliver beregnet 14 avlsværdital
 - Uden udenlandsk information
 - Med heterogen varians korrektion
- De 14 avlsværdital bliver genberegnet
 - Data prækorrigeret for heterogen varians
 - Med udenlandsk information

- De 14 avlsværdital bliver konverteret til de 36 parametre der indgår i formlerne for "avlsværditalkurverne"
- Sum af ydelse i 305 dage bliver beregnet separat for mælk, protein og fedt og separat for 1., 2. og 3. laktation, i alt 9 avlsværdital pr. dyr. For hver laktation bliver der derudover beregnet avlsværdital for fedt- og proteinprocent samt for persistens for mælkeydelse. Det giver yderligere 9 avlsværdital pr. dyr.
- Alle 18 avlsværdital bliver standardiseret med hensyn til gennemsnit og variation. Resultatet er 18 indekser med gennemsnit 100 og en spredning på 10 indekseheder
- Indekser for de enkelte laktationer bliver summeret med vægtene 0,50, 0,30 og 0,20 for henholdsvis 1., 2. og 3. laktation. Resultatet er 6 indekser: M-, P- og F-indeks, indeks for fedt- og proteinprocent samt persistensindeks.
- Endeligt bliver Y-indekset beregnet med de vægtforhold, som gælder for racegruppen.

Tabel 2. Genetiske parametre for RDM i den nordisk testdagsmodel – opgjort for 305-dages ydelser. Heritabiliteter på diagonalen, genetiske korrelationer over diagonalen, fænotypiske korrelationer under diagonalen

	1. laktation			2. laktation			3. laktation		
	Mælk	Protein	Fedt	Mælk	Protein	Fedt	Mælk	Protein	Fedt
Mælk1	0.41	0.86	0.62	0.91	0.79	0.53	0.87	0.73	0.45
Protein1	0.92	0.35	0.78	0.76	0.91	0.69	0.73	0.86	0.63
Fedt1	0.80	0.87	0.41	0.52	0.76	0.93	0.52	0.74	0.91
Mælk2	0.54	0.49	0.41	0.24	0.80	0.51	0.95	0.75	0.44
Protein2	0.49	0.52	0.48	0.93	0.21	0.77	0.77	0.97	0.72
Fedt2	0.42	0.47	0.55	0.84	0.90	0.28	0.52	0.75	0.99
Mælk3	0.48	0.44	0.37	0.53	0.51	0.44	0.20	0.80	0.48
Protein3	0.43	0.47	0.43	0.49	0.56	0.51	0.91	0.19	0.73
Fedt3	0.36	0.42	0.51	0.42	0.50	0.57	0.84	0.87	0.25

