

Til:	Projekt 4085, Kælvekvier, AP1b_PIN	Ansvarlig	JNI
		Oprettet	25-10-2017
Fra:	Jørgen Nielsen, HusdyrDigital	Side	1 af 6

Analyse af sammenhænge til besætningers andel af høje celletal ved første-kalvs-kørs første ydelseskontrol

Tillæg vedr. Udsætning og Malkesystem

Dette notat er et tillæg til mit notat version 1.1 af 03-10-2017. I dette notat bruges samme datasæt som tidligere, men der er tilføjet oplysninger om

- Hvorvidt en ko er udsat inden for 150 DEK (udsætning = død, aflivning eller slagtning)
- Malkesystem pr. 01-04-2016 (midt i dataperioden)
- For AMS også oplysninger om fabrikat

I de første analyser i dette notat bruges fortsat de to 'korrigerende' variable: Racegruppe og EKMSnit (ydelsesniveau). Begge variable er på besætningsniveau og er beskrevet i det tidligere notat.

Analyser

Jf. min forklaring i tidligere notat, undersøger jeg derfor en model med data på besætningsniveau:

Andel høje celletal
= EKMSnit + race + EKMSnit * race
+ **Gruppe**

Hvor **Gruppe** er enten:

- gruppeinddeling af besætninger på baggrund af udregnet andel at køer udsat inden for 150 DEK,
- Malkesystem,
- Fabrikat af AMS, eller
- Malkesystem, hvor AMS er opdelt i to fabrikater

Andel køer udsat inden for 150 DEK

Besætninger deles op i fire lige store grupper efter deres andel af udsatte køer:

Gruppe	Andel køer udsat inden for 150 DEK	Antal besætninger i gruppen	Besætningernes gennemsnitlige andel af høje celletal
1	Op til 3,0%	499	41,5%
2	3,0% – 5,3%	500	42,7%
3	5,3% – 8,7%	502	44,1%
4	8,7% og op	494	45,8%
	I alt	1.995	

I denne model er Andel udsatte køer signifikant. Ved nærmere undersøgelse ses, at der ikke er signifikant forskel på to grupper lige ved siden af hinanden, dvs. 1 og 2, samt 2 og 3 samt 3 og 4. Men der er signifikante forskelle, når vi sammenligner gruppe 1 med 3 eller 4 og når vi sammenligner gruppe 4 med 1 eller 2.

Malkesystem

Besætningerne er delt op efter hvilket malkesystem, som de stod registreret med pr. 1. april 2016 (hvilket var midt i perioden for hvor jeg har hentet data).

De fordeler sig således:

Malkesystem	Antal besætninger i gruppen	Besætningernes gennemsnitlige andel af høje celletal
AMS/VMS	511	51,2%
Bindestald	253	41,0%
Karrusel	129	
Side-By-Side	174	
Sildeben	910	
Tandem	18	
I alt	1.995	

Der er signifikant sammenhæng mellem malkesystem og andel høje celletal, men der er ikke signifikante forskelle, når vi sammenligner de malkesystemer, som ikke er AMS.

AMS-fabrikat

Langt hovedparten af de 511 besætninger med AMS har deres anlæg fra enten Lely eller DeLaval. Når vi sammenligner besætninger med disse to fabrikater fås følgende:

AMS-fabrikat	Antal besætninger i gruppen	Besætningernes gennemsnitlige andel af høje celletal
DeLaval	215	50,4%
Lely	288	46,3%
I alt	503	

Bemærk: man kan undre sig over, at procenterne i denne tabel ligger mellem 45 og 50, når procenten for AMS i tabellen ovenfor ligger på over 50. For der er jo tale om stort set de samme besætninger. Men dér skal man huske, at der i disse beregninger er korrigeret for effekten af besætningens race og ydelsesniveau. De procenttal, som er vist her, er såkaldte 'least square means' (LS-means). Og de er beregnet på denne måde bl.a. for at se *forskellene* mellem grupperne, når vi korrigerer for race og ydelse. Og den korrektion, som er forårsaget af race og ydelse, kan godt være forskellig i de to tabeller.

Forskellen mellem de to fabrikater er signifikant.

Malkesystem inkl. fabrikat

Man kan måske have lyst til at kombinere de to tabeller/modeller ovenfor, og lave en analyse, hvor vi kigger på malkesystem, men betragter AMS som værende to systemer. Jeg vælger at udelader gruppen af besætninger med Tandem-malkesystem, blandt andet fordi gruppen er ret lille. Dermed får vi følgende tabel:

Malkesystem	Antal besætninger i gruppen	Besætningernes gennemsnitlige andel af høje celletal
DeLaval (AMS)	215	53,1%
Lely (AMS)	288	49,5%
Bindestald	253	40,8%
Karrusel	129	
Side-By-Side	174	
Sildeben	910	
I alt	1.969	

I denne model er der signifikante forskelle mellem de to AMS-fabrikater, og de adskiller sig signifikant fra systemerne uden AMS. Men systemerne uden AMS adskiller sig ikke signifikant fra hinanden.

Egen ref: "U:\KvaegSASpc\DataGruppe\Projekter\KlvKvier2017\AP1b_PIN\ Datatræk3_Uds_MlkSys.sas.sas"

Analyser på dyr-niveau – udsætning ift. højt celletal

Ovenstående analyser er lavet på besætningsniveau, hvor vi kigger på sammenhængen mellem fx besætningens andel af udsatte 1. kalvs køer inden 150 DEK i forhold til andelen af høje celletal ved første ydelseskontrol.

Udvælgelse af model

I dette afsnit vil jeg forsøgsvis grave et spadestik dybere – her ved at kigge på dyr-niveau. Jeg vil ikke inddrage malkesystem i analyse – dels for at holde analysen relativt enkel og dels fordi jeg vil bruge en mixed model med besætning som random factor. Dermed vil en del af effekten af malkesystem være forklaret ved den tilfældige effekt af besætning, som vi lægger ind i modellen. Hvis vi sammenligner med analysen ovenfor, står vi altså med følgende variable:

- a. Højt celletal ved første ydelseskontrol (ja/nej for den enkelte ko)
- b. Udsat inden for 150 DEK (ja/nej for den enkelte ko)
- c. Besætningens ydelse
- d. Racen – enten dyrets race eller besætningsracen.
- e. Besætningen selv (random factor)

Jeg tænker at vende analysen sådan, at vi ønsker at forklare, om risikoen for udsætning hænger sammen med, om celletallet var højt ved første ydelseskontrol. Denne vinkel tænker jeg, er mere lige til i vores tolkning af resultaterne. For det vil så betyde, at vi med viden om det første celletal – får en indikation på risikoen for udsætning. Dvs. variable a og b.

Besætningens ydelse (variabel c) tænker jeg at udelade af analyse. Det gør jeg med samme argumentation som for malkesystem ovenfor.

I forhold til race vælger jeg at bruge oplysning om *dyrets* race (variabel d). Dels fordi jeg har den oplysning og fordi jeg laver analysen på dyr-niveau, og fordi jeg tænker, at det giver et nemmere tolkning af modellens resultater.

Besætningen selv (variabel e) skal indgå i analysen, for netop at gøre den til en mixed model analyse. På den måde tager jeg højde for at dyr, som er i samme besætning har en tendens til at ligne hinanden.

Min respons-variabel (b) er en ja/nej-variabel, og dermed har jeg noget som ligner en binomialfordeling. Dvs. vi skal over i en Generaliseret Lineær Model, og da besætningen er en random factor, bliver det i alt til en Generaliseret Lineær Mixed Model (GLMM). Det er så heldigt, at det har jeg fået undervisning i i Norge i uge 41, 2017.

Jeg tænker derfor på to modeller, som jeg vil estimere og teste. Begge modeller med udsætning (variabel b) som respons, og med besætning som random factor. Som forklarende variabel vælger jeg højt celletal (variabel a), og jeg kan så vælge at have dyrets race med eller ej. Skrevet op som mere matematisk notation eller R-kode, har vi følgende to modeller:

1. Udsætning ~ højt celletal + (besætning)
2. Udsætning ~ højt celletal + race + (besætning)

Analyseresultat

I begge modeller får vi signifikant virkning af højt celletal – og også af race i model 2. Det vil si-ge, at når vi følger det enkelte dyr, så er der en sammenhæng mellem celletal ved første ydel-seskontrol og risikoen for udsætning inden 150 DEK. Også når vi tager hensyn til racen er denne sammenhæng signifikant.

Model 1

Kigger vi på model 1 (Udsætning ~ højt celletal + (besætning)), så får vi som sagt en signifikant virkning af højt celletal. Ud fra det output som ses i appendiks fås følgende estimater:

Højt celletal	Estimat på logit-skala (A)	Estimeret sandsynlighed for udsætning af koen (B)	Empirisk beregnet andel udsatte køer (C)
Nej	-3.153192	4,1%	4,6%
Ja	-2.516360	7,5%	8,4%
B = inv.logit(A) C er beregnet ud fra rå data – er tænkt som en slags realitetstjek, men der er i denne beregning ikke taget hensyn til effekt af besætninger. Så den mest troværdige estimat anser jeg for at være B. Specielt hvis ikke det afviger absurd meget fra C.			

Dvs. for de køer, som *ikke* havde højt celletal, forventes en risiko for udsætning inden 150 DEK på 4,1%.

For de køer, som *havde* højt celletal ved første ydelseskontrol, forventes en risiko for udsætning inden 150 DEK på 7,5%.

Denne forskel er signifikant i analysen.

Dermed har vi altså næsten dobbelt så høj risiko for udsætning blandt de 1. kalvs køer, som har haft højt celletal ved første ydelseskontrol.

I analysen indgår 1.970 besætninger med over 20 kælvnings og i alt 154.464 kælvnings.

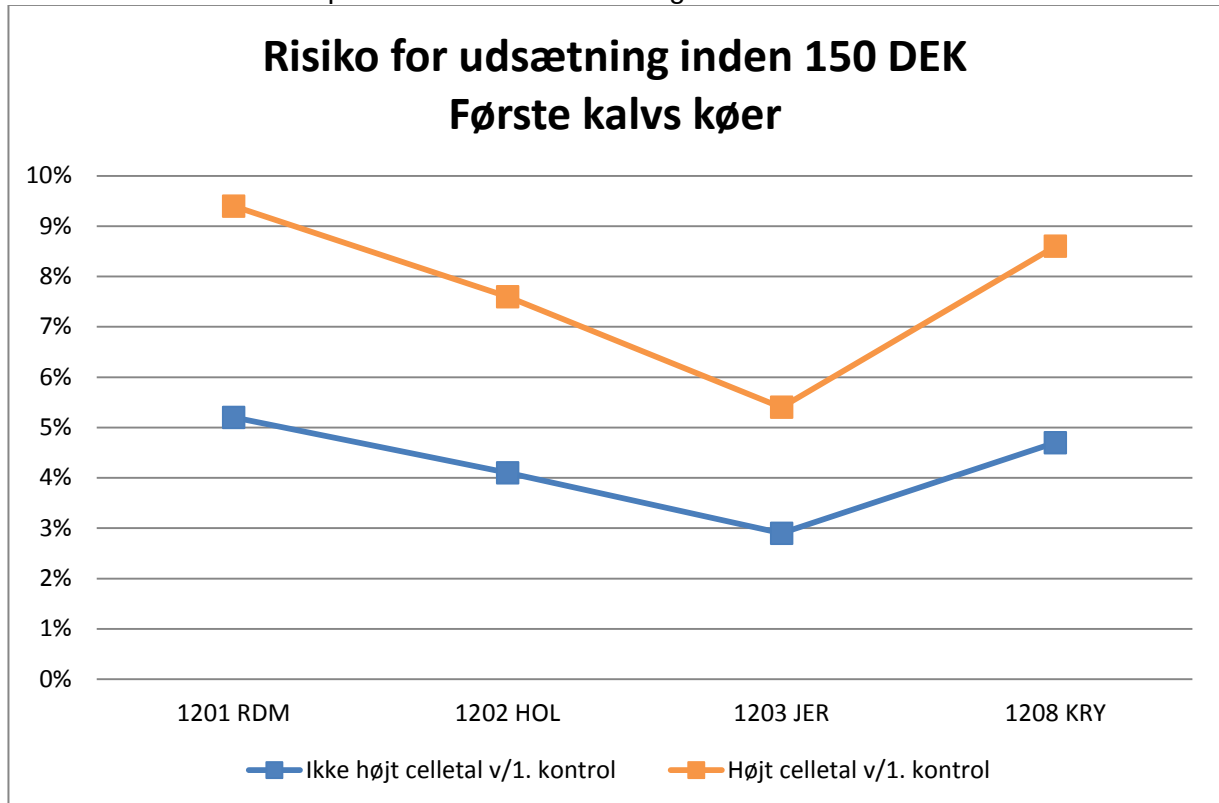
Model 2

Kigger vi på model 2 (Udsætning ~ højt celletal + race + (besætning)), så får vi også signifikant virkning. Både af højt celletal og af race. Det kan også tolkes som at vi får signifikant virkning af celletal – også når vi korrigerer for race. Ud fra output som ses i appendiks har vi følgende estimater:

Højt celletal	Race	Estimat på logit-skala (A)	Estimeret sandsynlighed for udsætning af koen (B)	Empirisk beregnet andel udsatte køer (C)
Nej	1201 RDM	-2.908435	5,2%	5,7%
	1202 HOL	-3.144921	4,1%	4,6%
	1203 JER	-3.502768	2,9%	3,5%
	1208 KRY	-3.005691	4,7%	5,6%
Ja	1201 RDM	-2.266943	9,4%	9,8%
	1202 HOL	-2.503429	7,6%	8,6%
	1203 JER	-2.861276	5,4%	6,0%
	1208 KRY	-2.364199	8,6%	9,3%
B = inv.logit(A) C er beregnet ud fra rå data – er tænkt som en slags realitetstjek, men der er i denne beregning ikke taget hensyn til effekt af besætninger. Så den mest troværdige estimat anser jeg for at være B. Specielt hvis ikke det afviger absurd meget fra C.				

I analysen indgår 1.962 besætninger med over 20 kælvnings og i alt 153.022 kælvnings.

Grafisk kan resultaterne præsenteres som i denne graf:



Igen ses en næsten fordobling af risikoen for udsætning, hvis celletallet er højt ved første ydelseskontrol.

Egen ref: "U:\KvaegSASpc\DataGruppe\Projekter\KlvKvier2017\AP1b_PIN\Analyse_Udsaet_Hicel_1.R"

Appendiks: output fra R vedr. analyse på dyr-niveau

Summary af model 1: Udsætning ~ højt celletal + (besætning)

```
Family: binomial ( logit )
Formula:          udsat ~ fhicel + (1 | fchnr)
Data: cel2

      AIC      BIC   logLik deviance df.resid
70228.3 70258.2 -35111.2  70222.3   154461

Random effects:

Conditional model:
Groups Name      Variance Std.Dev.
fchnr (Intercept) 0.3182   0.5641
Number of obs: 154464, groups: fchnr, 1970

Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.15319     0.02211 -142.61  <2e-16 ***
fhicel1      0.63683     0.02180   29.21  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary af model 2: Udsætning ~ højt celletal + race + (besætning)

```
Family: binomial ( logit )
Formula:          udsat ~ fhicel + frace_id + (1 | fchnr)
Data: celR2

      AIC      BIC   logLik deviance df.resid
69510.5 69570.2 -34749.3  69498.5   153016

Random effects:

Conditional model:
Groups Name      Variance Std.Dev.
fchnr (Intercept) 0.3028   0.5502
Number of obs: 153022, groups: fchnr, 1962

Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.90844     0.05436 -53.51  < 2e-16 ***
fhicel1      0.64149     0.02192   29.27  < 2e-16 ***
frace_id1202 -0.23649     0.05499  -4.30 1.70e-05 ***
frace_id1203 -0.59433     0.07363  -8.07 6.89e-16 ***
frace_id1208 -0.09726     0.06239  -1.56  0.119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```