

Statistisk vurdering af opgørelse af læssetid og slagsår

David Edwards,
Det Jordbrugsvidenskabelige Fakultet,
Aarhus Universitet

June 9, 2010

1 Baggrund

Denne rapport er en statistisk vurdering af dokumentet "Analyse af læssetid for slagtesvin i forbindelse med politianmeldelse af slagsår" forfattet af Niels Peter Baadsgaard, Videnscenter for Svineproduktion, dateret 11. januar 2010. Dokumentet er i det følgende refereret til som NPB. Formålet med analysen var at undersøge, om der er evidens for en sammenhæng mellem politianmeldelser af slagsår og læssetid per gris, idet man kunne forvente at et nyt afregningssystem kunne tilskynde transportøren til at slå grisene for at læsse dem hurtigere. NPB fandt en meget beskedent forskel mellem median læssetid per gris for de leverancer hvor der skete politianmeldelse, i sammenligning med en kontrolgruppe af leverancer hvor der ikke skete anmeldelser, og konkluderede, at læssetiden per gris i forbindelse med politianmeldelser ikke adskiller sig fra læssetider per gris iverdigt. En del observationer måtte dog udelades fra kontrolgruppen på grund af registreringsfejl, hvilket kunne tænkes at svække sammenligningsgrundlaget.

Formålet med nærværende rapport er at vurdere den anvendte analyse af materialet, herunder

- at redegøre for frasorteringen i dannelsen af kontrolgruppen og vurdere om dette kunne føre til bias i konklusionerne,
- at vurdere om andre analysemetoder til belysning af en eventuel sammenhæng mellem læssetider per gris og politianmeldelse bør anvendes,

På baggrund af de nævnte gennemsnitlige leverance størrelser og samlet leveringstider i NPB's redegørelse og den tilsyneladende forskel i disse belyses også

- sammenhæng mellem leverance størrelse og politianmeldelse, og
- sammenhæng mellem leverancernes samlet læssetid og politianmeldelse.

skønt dette ikke indgik i den oprindelige opgave stillet af FVST.

2 Resultater

2.1 Datagrundlaget

Den nærværende analyse er baseret på tre datafiler, tilsendt fra Niels Peter Baadsgaard. Det første datasæt indeholder data om 141385 leverancer i perioden 2006-2008, og som ikke er tilknyttet politianmeldelser. Det andet datasæt er det reducerede datasæt, hvor problematiske records er blevet frasorteret. Dette indeholder data om 69621 leverancer. Det tredje datasæt indeholder data om 52 leverancer som førte til politianmeldelser. I alt var der 100 politianmeldelser i perioden 2006-2008, men 48 er korrekt blevet frasorteret på grund af registreringsfejl eller manglende oplysninger.

Datafilerne indeholder bl.a. information om dato, start tidspunktet, leverance størrelse (antal grise), læsenr og læssetid for leverancen. I datasættet uden frasortering er der et stort antal fejlregistreringer, som kan illustreres ved:

	dato	leverede	laesnr	tidperlev	Starttid
57	2008-06-10	3	900	4	06:46:00
58	2008-06-10	3	2034	1	11:34:00
59	2008-06-10	3	900	1	06:46:00
739	2008-03-17	265	1400	41	12:13:00
740	2008-03-17	265	1930	69	17:36:00
745	2008-06-11	139	3032	56	09:08:00
746	2008-06-11	139	3020	37	07:56:00
747	2008-06-11	139	3020	4	07:56:00
748	2008-06-11	108	3032	56	09:08:00
749	2008-06-11	108	3020	37	07:56:00
750	2008-06-11	108	3020	4	07:56:00

Det fremgår at en leverance kan omfatte flere læs, men at antal grise i disse er ikke registreret. Desuden er der hyppigt dobbeltregistreringer, ofte med forskellige læsnumre. Andre fejl forekommer også. I NPB er der valgt at frasortere alle leverancer hvor der ikke er entydig identifikation af dato, chr og leverandørnummer, hvor der er flere læs per leverance, og hvor læssetid er anført som nul.

Analysen i NPB er primært fokuseret på evt sammenhænge mellem forekomst af politianmeldelse og læssetid per gris. Læssetid per gris er beregnet ud fra den total læssetid og leverance størrelse.

2.2 Gruppesammenligninger

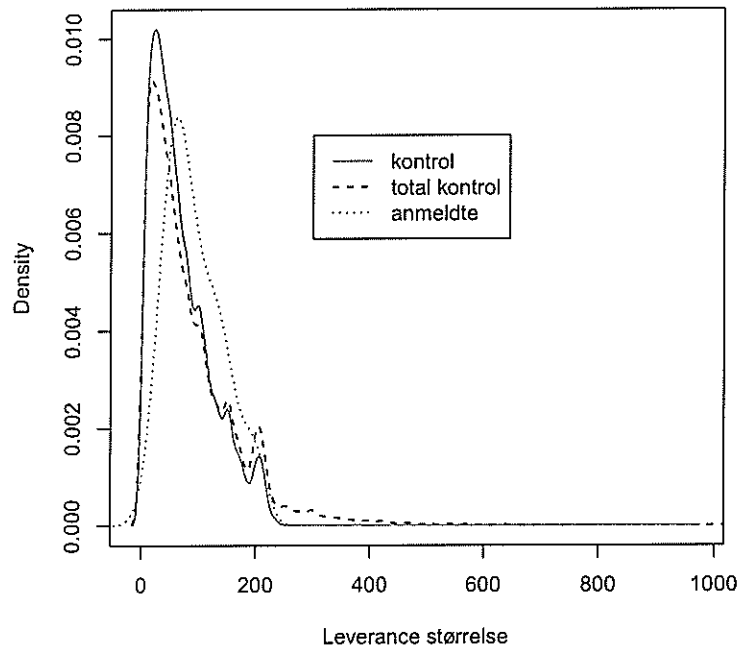
I det følgende sammenlignes de tre datafiler, først med hensyn til fordeling af leverance størrelse. Kvantilerne er her vist:

```
[1] "Kvantiler af leverance størrelse"
```

	0%	25%	50%	75%	100%
total kontrol	1	29	63	121	1465
kontrol	1	28	55	100	961
anmeldte	12	57	79	125	203

Det fremgår at frasorteringen har reduceret medianen for leverance størrelsen fra 63 til 55. For de leverancer der fører til politianmeldelser, er median 79. At den maksimale leverance størrelse er noget mindre blandt de anmeldte leverancer kan tilskrives de langt større antal data i de øvrige datasæt.

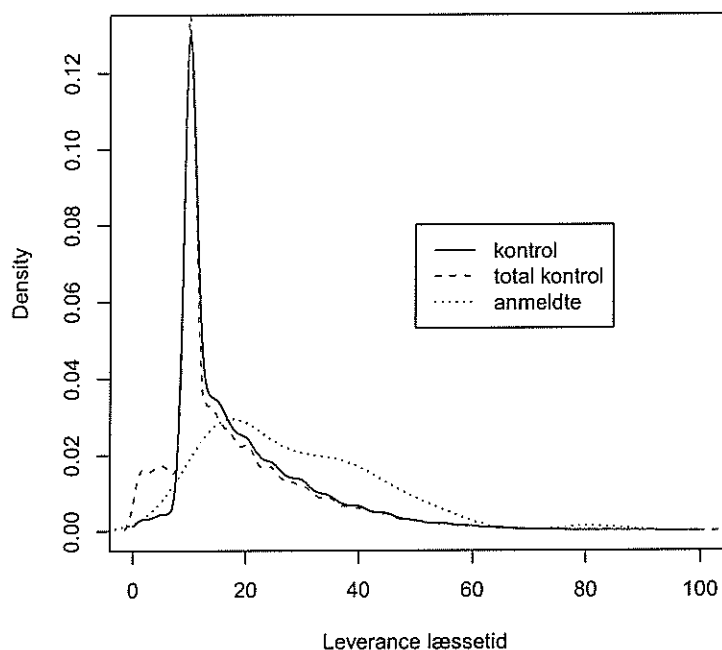
Grafisk fremstillet kan de tre fordelinger illustreres ved:



Tilsvarende tabel og graf over fordeling af leverancernes læssetid vises her:

[1] "Kvantiler af leverance læssetid"

	0%	25%	50%	75%	100%
total kontrol	1	10	13	23	387
kontrol	1	10	15	25	234
anmeldte	10	17	26	37	81



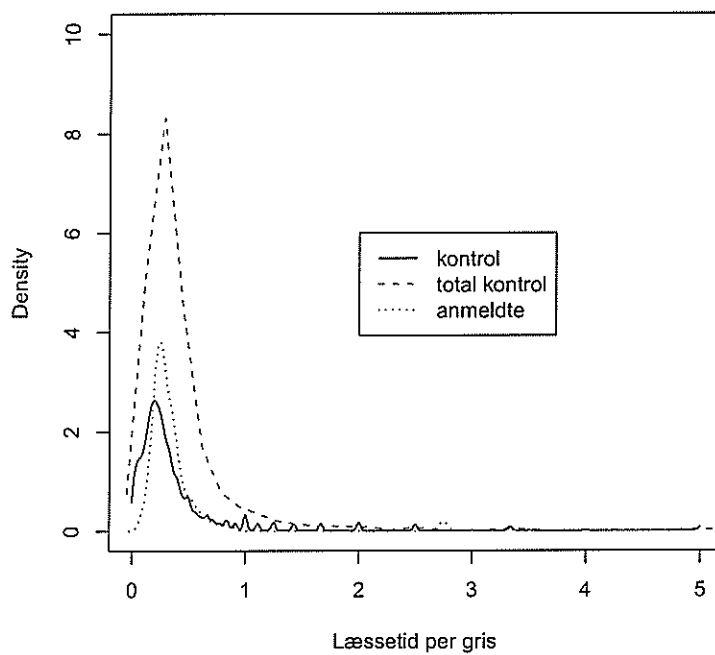
Det fremgår at median læssetiden for de anmeldte leverancer er 26 min, som er noget højere end i den frasorterede kontrol gruppe, hvor medianen er 15 min. Det ses, at læssetiden er mindre påvirket af frasorteringen end leverance størrelsen. Fordelinger i det oprindelige og det reducerede datasæt er ganske ens, bortset fra de korte læssetider (fra 0-8 min).

Tilsvarende tabel over fordeling af læssetid per gris vises her:

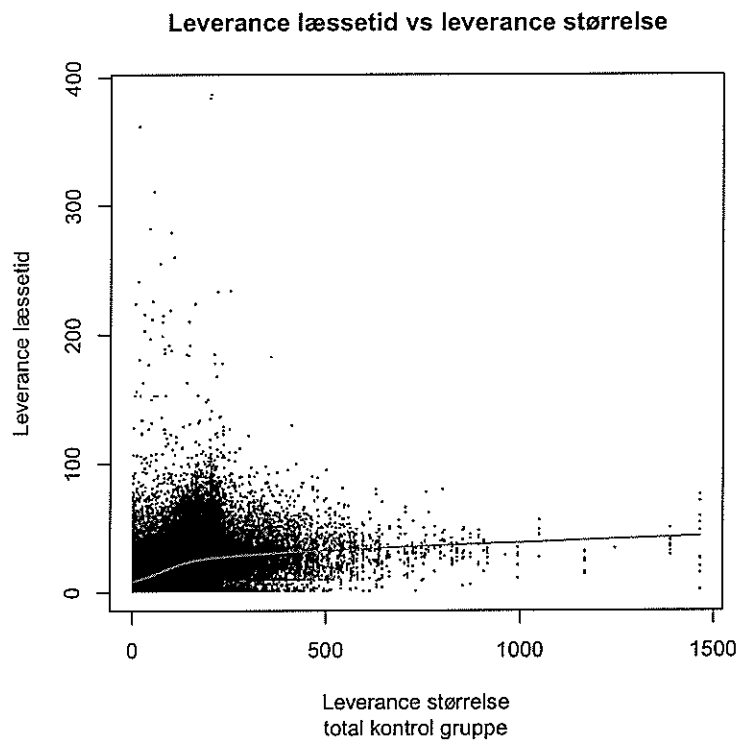
```
[1] "Kvantiler af læssetid per gris"

          0%  25%  50%  75% 100%
total kontrol 0.0014 0.16 0.26 0.42 91.0
kontrol      0.0050 0.21 0.29 0.45 85.0
anmeldte     0.1020 0.23 0.28 0.37  2.8
```

Som NPB bemærker, adskiller median læssetid pr gris blandt de anmeldte leverancer sig ikke væsentligt fra medianen i kontrolgruppen.

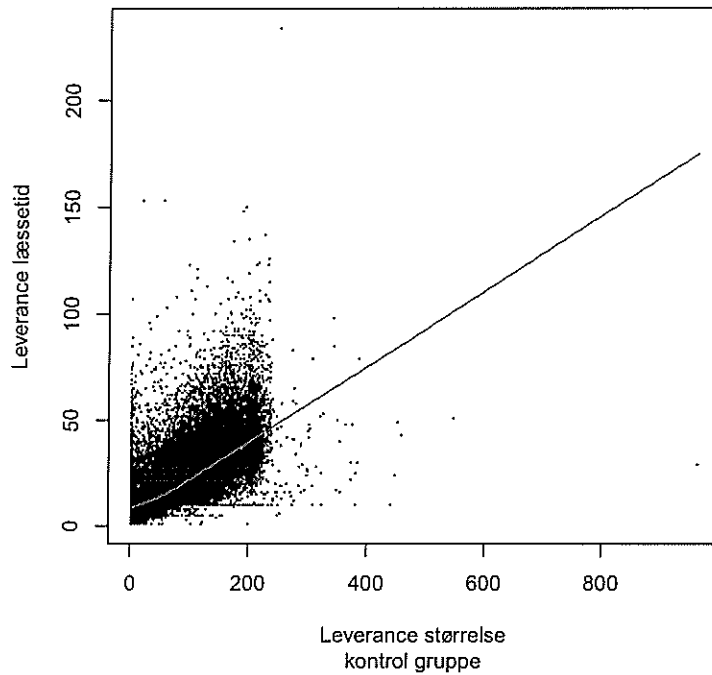


Yderligere indsigt i effekten af frasorteringen kan ses ved at sammenligne en graf, der viser leverance størrelse og læssetid i det oprindelige datasæt:



med den tilsvarende graf i det reducerede datasæt:

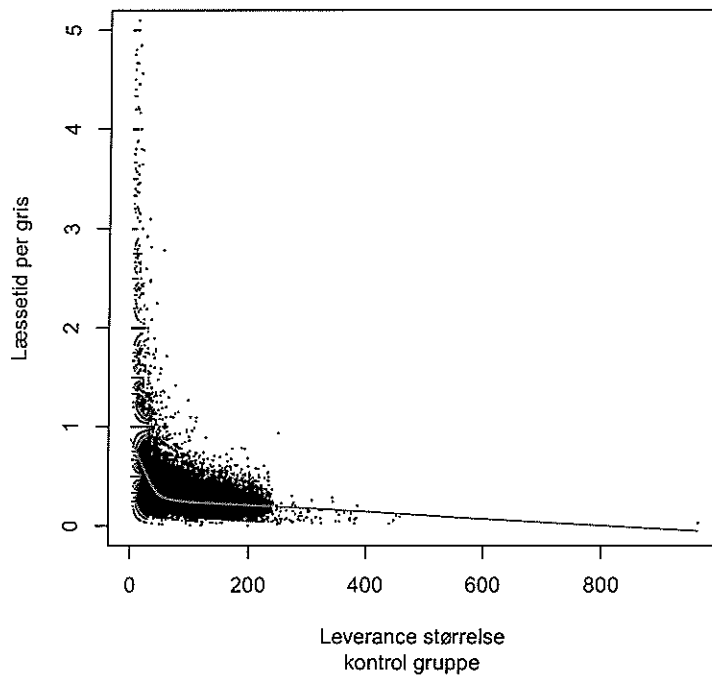
Leverance læssetid vs leverance størrelse



Graferne viser også en estimeret kurv (lowess smoothing) for forholdet mellem læssetid og antal gris. For det reducerede datasæt er forholdet linært, som forventet.

Tilsvarende graf af forholdet mellem læssetid per gris og leverance størrelse følger:

Læssetid per gris vs leverance størrelse



Det ses, at læssetiden per gris er en faldende funktion af leverance størrelsen, hvilket antageligt afspejler større effektivitet ved de større læs.

2.3 Analyse af sammenhængen mellem politianmeldelse, læssetid per gris, leverance størrelse og samlet leverance læssetid

Som en mere formel statistisk analyse af eventuelle sammenhænge mellem forekomsten af politianmeldelser og leverance størrelse, læssetid per leverance, og læssetid per gris, er der udført en logistisk regressionsanalyse, med gruppe (dvs, anmeldelse eller ingen anmeldelse) som respons variabel, og leverance størrelse, læssetid per leverance og læssetid per gris som forklarende variable. Det reducerede datasæt er anvendt. Resultaterne er som følger:

\$coefficients

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.6921803	0.3596239	21.38951	1.6729e-101
tidpergris	0.3989740	0.5058164	0.78877	4.3025e-01
antalgrise	0.0007794	0.0036745	0.21211	8.3202e-01
tidperlev	-0.0317213	0.0089758	-3.53408	4.0920e-04

Der er ingen evidens af en effekt af læssetid per gris ($P=0.4302$) eller leverance størrelse ($P=0.8320$) på responsvariablen, men der er klar evidens af en effekt af læssetid per leverance ($P=0.0004$). Ifølge analysen, desto større læssetid per leverance, desto større er sandsynligheden for politianmeldelse.

3 Konklusion

NPB konkluderede korrekt, at der på basis af en direkte sammenligning af median læssetid per gris blandt de anmeldte og de resterende registreringer ikke er evidens af en sammenhæng mellem forekomst af politianmeldelse og læssetid, opgjort som læssetid per gris. Ovenstående analyse antyder, at der derimod er en sammenhæng mellem anmeldelse og leverancens læssetid, når man tager hensyn til sammenhængen mellem læssetid og antal grise. Slagskader synes overvejende at være knyttet til læsninger af længere varighed.

En betydelig andel af registreringerne var nødvendigvis udeladt fra analysen på grund af fejlregistreringer af diverse typer. Der er derfor risiko for, at dette giver bias i sammenligningsgrundlaget. Imidlertid synes effekten af frasortering at være væsentligt mindre på leverancens læssetid end på leverancens størrelse. Det vurderes derfor, at der ikke er tale om væsentlig bias på leverancens læssetid.