

Kørselsafstand

Fremgangsmåde og resultater

Det af kunden udleverede datasæt indeholder information for 4165 bedrifter. For hver af disse er fugleflugtsafstanden (i km) til biogasanlæg (Halkjærvej 19, 6973 Ørnhøj) angivet.

Fugleflugtsafstandene er inddelt i de 10 afstandsklasser:

(0.07,12.1], (12.1,16.6], (16.6,19.7], (19.7,22.1], (22.1,24.9],
(24.9,27.4], (27.4,30.7], (30.7,34.2], (34.2,37.1], (37.1,40].

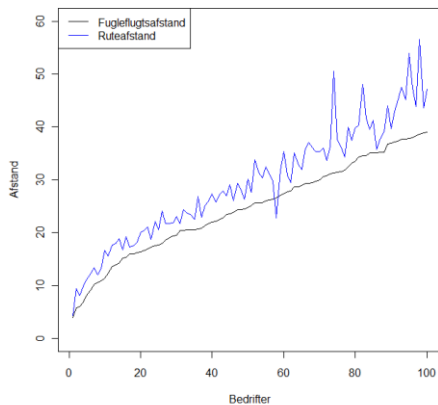
Der er udvalgt 100 bedrifter, således at hver afstandsklasse indeholder 10 tilfældige bedrifter.

For hver af disse 100 bedrifter er den korteste ruteafstand (i km) til biogasanlæg bestemt vha. KRAK ruteplan.

Det resulterende datasæt ses i tabel 1 på side 13.

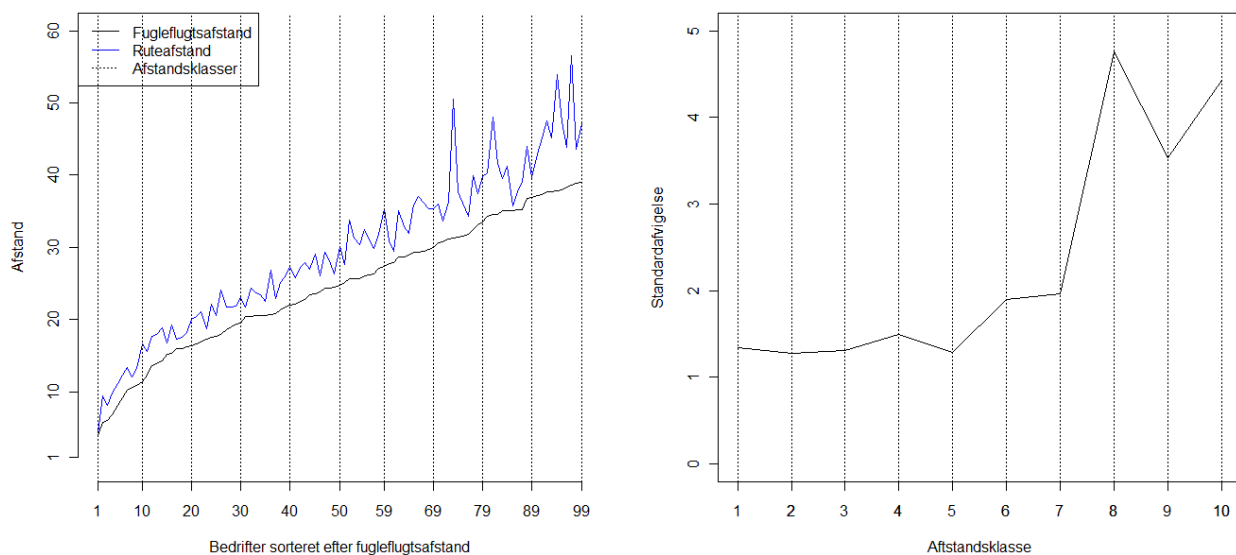
Et indledende plot (figur 1) viser fugleflugtsafstanden og ruteafstanden for de 100 bedrifter.

Figur 1: Plot af fugleflugtsafstand og ruteafstand.



Der er tilsyneladende en fejl-observation. Denne observation (observation 21 tilhørende afstandsklasse 6: "(24.9,27.4]" jf. tabel 1) fjernes fra datasættet, og plottet laves igen (til venstre i figur 2). Det ses, at ruteafstanden viger mere væk fra fugleflugtsafstanden, jo større afstanden er. Desuden er standardafvigelsen af forskellen mellem fugleflugt og ruteafstand inden for hver klasse plottet (til højre i figur 2).

Figur 2: T.v.: Plot af fugleflugtsafstand og ruteafstand efter fejl-observation er fjernet. T.h.: Standardafvigelsen af forskellen mellem fugleflugt og ruteafstand inden for hver afstandsklasse.

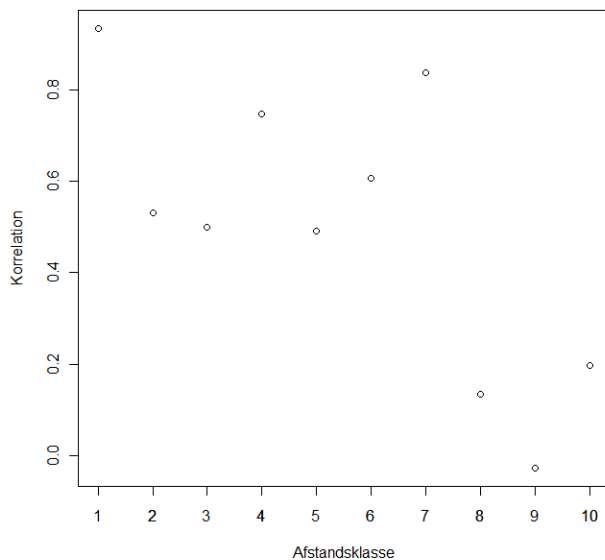


For hver afstandsklasse er korrelationen mellem fugleflugtsafstand og den korteste ruteafstand bestemt. Resultatet ses i nedenstående tabel 2 samt figur 3. Afstandsklasserne er i det følgende angivet som 1, ..., 10.

Tabel 2: Korrelationen for de forskellige afstandsklasser.

Afstandsklasse	cor(fugleFlugt, ruteAfstand)
1: (0.07, 12.1]	0.93531479
2: (12.1, 16.6]	0.53120067
3: (16.6, 19.7]	0.49900079
4: (19.7, 22.1]	0.74716837
5: (22.1, 24.9]	0.49190869
6: (24.9, 27.4]	0.60643204
7: (27.4, 30.7]	0.83850893
8: (30.7, 34.2]	0.13418716
9: (34.2, 37.1]	-0.02808370
10: (37.1, 40]	0.19812840

Figur 3: Korrelationen for de forskellige afstandsklasser.



Funktionen "korrelation = $f(\text{fugleflugtsafstand})$ " estimeres. Dette er udført på to forskellige måder:

1. Ikke-lineær regression med logistisk vækstkurve (side 5)
2. Step-funktion (side 7)

1. Ikke-lineær regression med logistisk vækstkurve

Det kan ikke umiddelbart lade sig gøre at fitte en logistisk vækstkurve til data. Pga. en generel forventning om at korrelationen falder med afstandsklassen, fjernes således afstandsklasse 7, som har en højere korrelationsværdi end forventet.

Modellen, der fittes til data er:

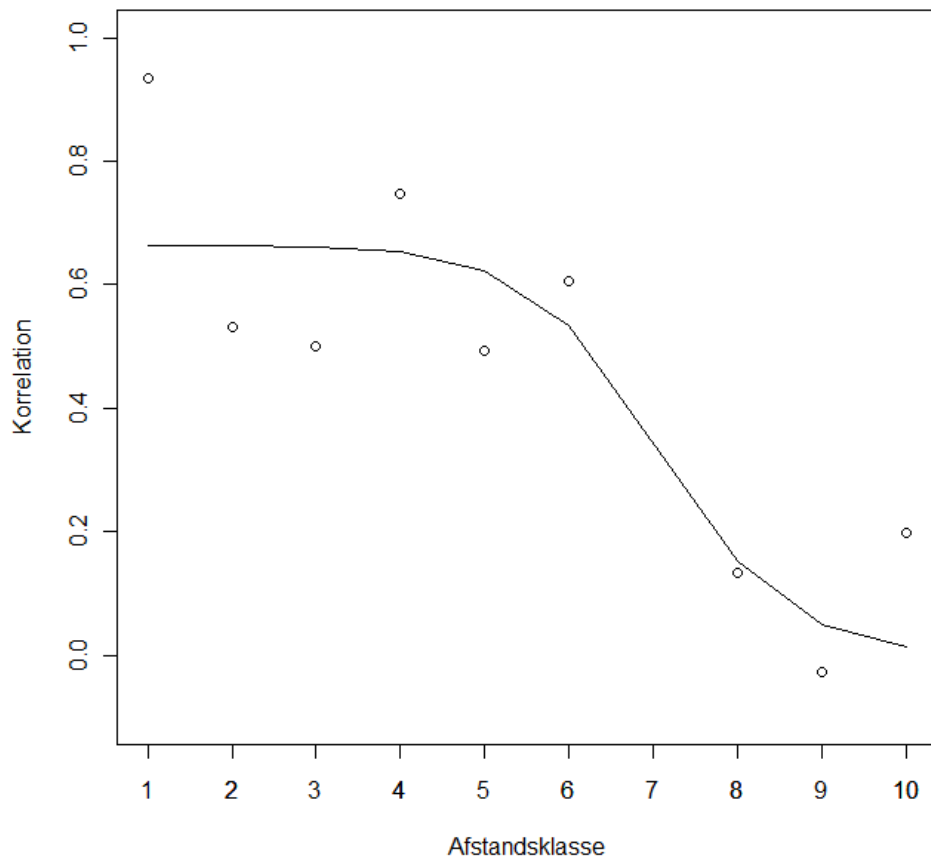
$$\text{korrelation} = \text{Asym} / (1 + \exp(-(\text{afstandsKlasse} - \text{xmid}) / \text{scal}))$$

De estimerede koefficienter og et tilhørende plot ses i tabel 3 og figur 4.

Tabel 3: Output fra R-kald

```
-----  
Nonlinear regression model  
Formula: corrVectorNy ~ logist(afstandsKlasserNy, Asym, xmid, scal)  
  
Parameters:  
      Estimate Std. Error t value Pr(>|t|)  
Asym  0.66397    0.09297   7.141  0.00038 ***  
xmid  7.07567    0.91138   7.764  0.00024 ***  
scal -0.75882    0.56739  -1.337  0.22956  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.1775 on 6 degrees of freedom  
Number of iterations to convergence: 29  
Achieved convergence tolerance: 8.103e-06  
-----
```

Figur 4: Den fittede logistiske vækstkurve og de tilhørende observationer.

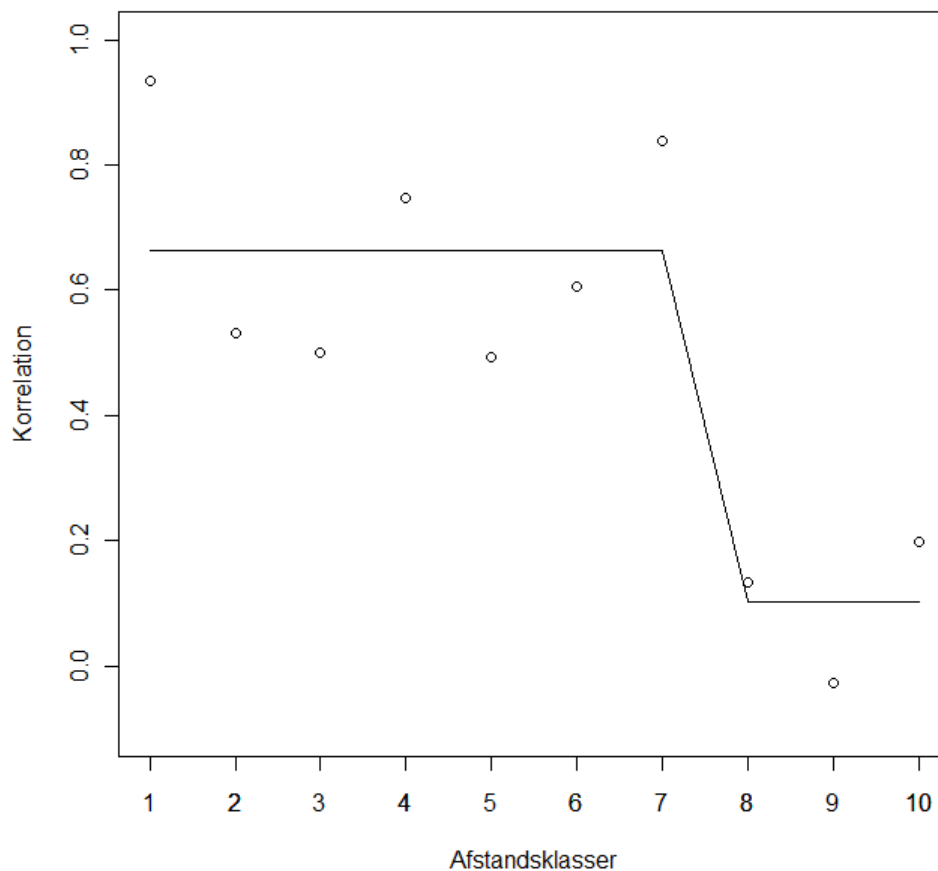


2. Step-funktion

Fra en betragtning af figur 3, fittes der en step-funktion til data. De fundne korrelationer ses at kunne adskilles i to grupper; afstandsklasserne 1 til 7, samt afstandsklasserne 8 til 10.

Middelkorrelationen i gruppen bestående af afstandsklasserne 1 til 7 er 0.6642192, og middelkorrelationen i gruppen bestående af afstandsklasserne 8 til 10 er 0.1014106.

Figur 5: Den fittede step-funktion og de tilhørende observationer.

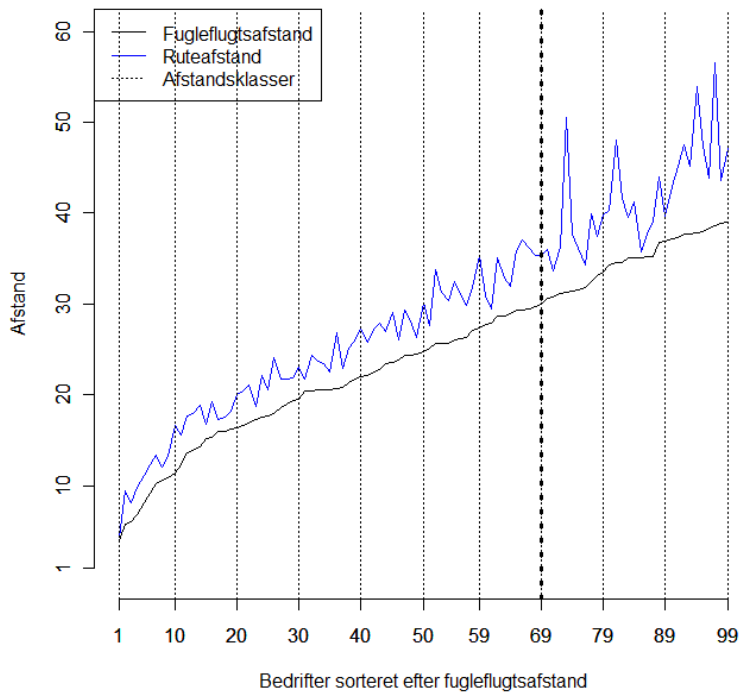


Fittet af den logistiske vækstkurve til data viste, at den asymptotiske korrelationsværdi er 0.66397, og at vendepunktet er ved 7.07567. Dette sammenholdt med figur 2 viser, at det er rimeligt at opdele data i de to grupper:

1. Afstandsklasse 1-7 (op til 30 km. fugleflugtsafstand)
2. Afstandsklasse 8-10 (mere end 30 km. fugleflugtsafstand).

For at kunne sige noget om hvor meget fugleflugtsafstanden bør øges for at få ruteafstanden, betragtes middeldifferencen mellem de to afstande for hver af de to grupper.

Figur 6: Plot af fugleflugtsafstand og ruteafstand, opdelt i afstandsklasser.



1. Afstandsklasse 1-7:

Middelforskel på fugleflugtsafstand og ruteafstand: 3.87 km.
Standard error: 0.2088429.

2. Afstandsklasse 8-10:

Middelforskel på fugleflugtsafstand og ruteafstand: 7.03 km.
Standard error: 0.8123164.

Det undersøges hvilke bedrifter, der skiller sig ud på plottet i figur 1. Figur 7 er tilsvarende figur 1, blot med disse bedrifter identificeret.

I det følgende er forskel i afstand beregnes som:
ruteafstanden - fugleflugtsafstanden.

For hver bedrift er der angivet den observerede afstandsforskel (i parentes), samt en web-adresse, der viser ruten på Krak.

De tre bedrifter med den mindste forskel i afstand er:

Bedrift 21: Røgindvej 18, 6940 Lem St.	(-3.84 km)
http://map.krak.dk/m/IVvBY	
Bedrift 28: Holmgårdsvej 2, 6971 Spjald	(0.30 km)
http://map.krak.dk/m/IVvBg	
Bedrift 40: Vardevej 54, 7280 Sønder Felding	(0.55 km)
http://map.krak.dk/m/IVvAV	

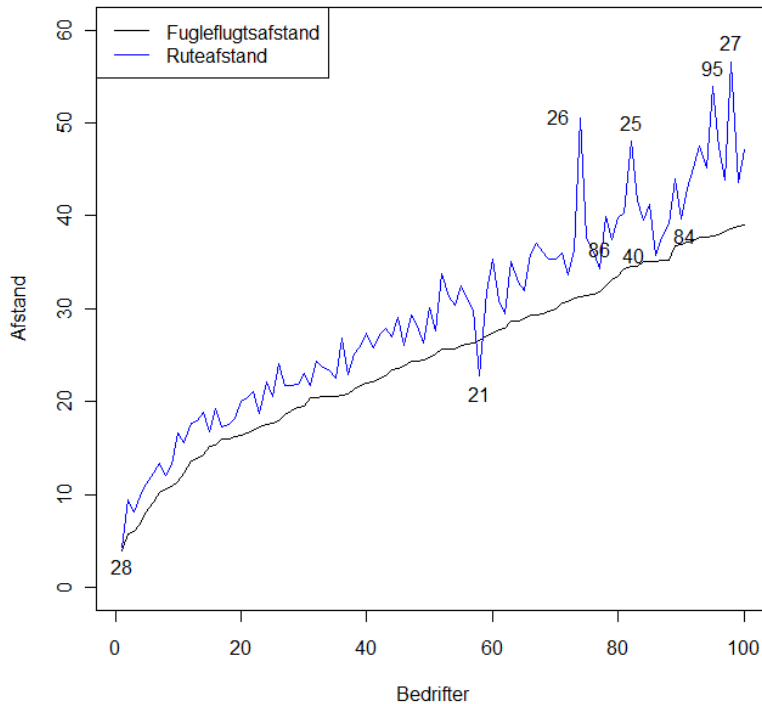
De fire bedrifter med den største forskel i afstand er:

Bedrift 25: Minervavej 31, 6960 Hvide Sande	(13.37 km)
http://map.krak.dk/m/IVZYw	
Bedrift 95: Tambohusvej 17, 7790 Thyholm	(16.08 km)
http://map.krak.dk/m/IVZpp	
Bedrift 27: Sønder Klitvej 97, 6960 Hvide Sande	(18.01 km)
http://map.krak.dk/m/IVZI9	
Bedrift 26: Solbakken 7, 6960 Hvide Sande	(19.25 km)
http://map.krak.dk/m/IVZ9P	

De tre bedrifter med den mindste forskel i afstand indenfor afstandsklasserne 8-10 er:

Bedrift 40: Vardevej 54, 7280 Sønder Felding	(0.55 km)
http://map.krak.dk/m/IVZba	
Bedrift 84: Lemvigvej 6, 7600 Struer	(2.48 km)
http://map.krak.dk/m/IVZJ7	
Bedrift 86: Stadion Alle 1, 7600 Struer	(2.51 km)
http://map.krak.dk/m/IVZ0U	

Figur 7: Plot af fugleflugtsafstand og ruteafstand med forskellige interessante bedrifter identificeret.



Bedrifterne 25, 26 og 27 ligger alle i Hvide Sande, hvilket medfører en stor forskel i fugleflugtsafstand og ruteafstand. Disse observationer tages ud af datasættet. Umiddelbart ser det fra Krak også underligt ud, at der er så stor forskel i flugleflugt og ruteafstand for bedrift 95. Denne fjernes også fra datasættet.

Som før tages bedrift 21 ud af datasættet, da dette antages at være en fejlobservation. Det samme gør sig gældende for bedrift 40, hvor der ved nærmere eftersyn viser sig at være en længere ruteafstand end indtastet.

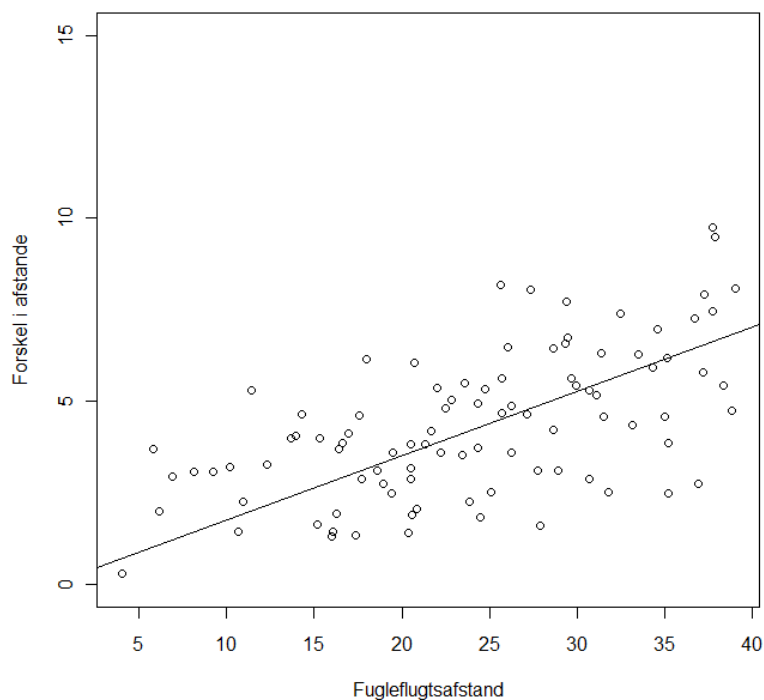
Efter disse bedrifter er fjernet fra datasættet er den observerede forskel i afstand plottet som funktion af fugleflugtsafstanden. Til disse datapunkter er der forsøgt at fitte en eksponentiel vækstkurve og en lineær sammenhæng. Resultaterne ses i figur 8 og tabel 4, samt figur 9 og tabel 5 nedenfor.

Fittet med den lineære sammenhæng vurderes ud fra de beregnede "Root Mean Squared Error" at beskrive data bedst, og fra tabel 4 ses at følgende sammenhæng er fundet:

$$\text{forsk} = 0.176033 * \text{fugleflugtsafstand},$$

hvor koefficienten til fugleflugtsafstand har standard error givet ved 0.006669.

Figur 8: Forskellene mellem ruteafstand og fugleflugtsafstand plottet op mod fugleflugtsafstanden. Det lineære fit er indtegnet.



Tabel 4: Output fra R-kald.

Call:

```
lm(formula = forskellene ~ fuglene - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7309	-0.9599	0.2029	1.3545	3.6818

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
fuglene	0.176033	0.006669	26.40	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

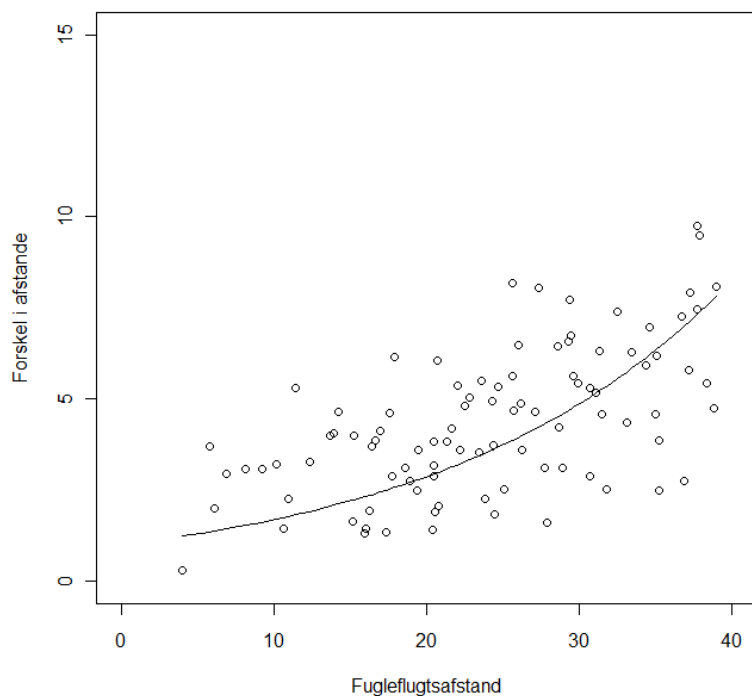
Residual standard error: 1.654 on 93 degrees of freedom

Multiple R-squared: 0.8822, Adjusted R-squared: 0.881

F-statistic: 696.8 on 1 and 93 DF, p-value: < 2.2e-16

Beregnet Root Mean Squared Error: 1.645069

Figur 9: Forskellene mellem ruteafstand og fugleflugtsafstand plottet op mod fugleflugtsafstanden. Det eksponentielle fit ($e^{(\text{fugleflugt} \cdot a)}$) er indtegnet.



Tabel 5: Output fra R-kald.

Formula: forskellene ~ expo(fuglene, a)

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
a	0.052740	0.001323	39.87	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.75 on 93 degrees of freedom

Number of iterations to convergence: 4

Achieved convergence tolerance: 9.34e-06

Beregnet Root Mean Squared Error: 1.740518

Tabel 1: Det datasæt, som beregningerne er baseret på.
 Afstande er angivet i km.

	fugleFlugt	ruteAfstand	Afstandsklasse
1	36.93	39.7	(34.2,37.1]
2	36.73	44	(34.2,37.1]
3	33.5	39.8	(30.7,34.2]
4	38.35	43.8	(37.1,40]
5	20.83	22.9	(19.7,22.1]
6	27.88	29.5	(27.4,30.7]
7	24.36	28.1	(22.1,24.9]
8	17.59	22.2	(16.6,19.7]
9	20.39	21.8	(19.7,22.1]
10	15.99	17.3	(12.1,16.6]
11	17.34	18.7	(16.6,19.7]
12	19.41	21.9	(16.6,19.7]
13	23.45	27	(22.1,24.9]
14	16.41	20.1	(12.1,16.6]
15	10.18	13.4	(0.07,12.1]
16	26.29	29.9	(24.9,27.4]
17	29.67	35.3	(27.4,30.7]
18	29.97	35.4	(27.4,30.7]
19	27.34	35.4	(24.9,27.4]
20	26.23	31.1	(24.9,27.4]
21	26.64	22.8	(24.9,27.4]
22	20.58	22.5	(19.7,22.1]
23	15.17	16.8	(12.1,16.6]
24	21.69	25.9	(19.7,22.1]
25	34.63	48	(34.2,37.1]
26	31.25	50.5	(30.7,34.2]
27	38.59	56.6	(37.1,40]
28	4	4.3	(0.07,12.1]
29	10.65	12.1	(0.07,12.1]
30	6.11	8.1	(0.07,12.1]
31	16.27	18.2	(12.1,16.6]
32	18.95	21.7	(16.6,19.7]
33	20.51	23.7	(19.7,22.1]
34	16.04	17.5	(12.1,16.6]
35	20.51	23.4	(19.7,22.1]
36	21.35	25.2	(19.7,22.1]
37	23.83	26.1	(22.1,24.9]
38	25.61	33.8	(24.9,27.4]
39	35.03	39.6	(34.2,37.1]
40	35.15	35.7	(34.2,37.1]
41	22.19	25.8	(22.1,24.9]

42	28.9	32	(27.4, 30.7]
43	28.66	32.9	(27.4, 30.7]
44	16.97	21.1	(16.6, 19.7]
45	31.52	36.1	(30.7, 34.2]
46	38.85	43.6	(37.1, 40]
47	19.48	23.1	(16.6, 19.7]
48	33.15	37.5	(30.7, 34.2]
49	15.29	19.3	(12.1, 16.6]
50	35.23	39.1	(34.2, 37.1]
51	35.12	41.3	(34.2, 37.1]
52	37.29	45.2	(37.1, 40]
53	29.31	35.9	(27.4, 30.7]
54	26.02	32.5	(24.9, 27.4]
55	29.38	37.1	(27.4, 30.7]
56	28.64	35.1	(27.4, 30.7]
57	37.75	47.5	(37.1, 40]
58	14.26	18.9	(12.1, 16.6]
59	8.12	11.2	(0.07, 12.1]
60	18.58	21.7	(16.6, 19.7]
61	6.86	9.8	(0.07, 12.1]
62	5.79	9.5	(0.07, 12.1]
63	13.95	18	(12.1, 16.6]
64	13.69	17.7	(12.1, 16.6]
65	11.41	16.7	(0.07, 12.1]
66	17.95	24.1	(16.6, 19.7]
67	22.02	27.4	(19.7, 22.1]
68	17.72	20.6	(16.6, 19.7]
69	16.64	20.5	(16.6, 19.7]
70	23.59	29.1	(22.1, 24.9]
71	10.94	13.2	(0.07, 12.1]
72	20.48	24.3	(19.7, 22.1]
73	22.5	27.3	(22.1, 24.9]
74	22.84	27.9	(22.1, 24.9]
75	25.68	31.3	(24.9, 27.4]
76	34.63	41.6	(34.2, 37.1]
77	12.33	15.6	(12.1, 16.6]
78	9.22	12.3	(0.07, 12.1]
79	24.47	26.3	(22.1, 24.9]
80	25.09	27.6	(24.9, 27.4]
81	20.73	26.8	(19.7, 22.1]
82	31.14	36.3	(30.7, 34.2]
83	24.75	30.1	(22.1, 24.9]
84	35.22	37.7	(34.2, 37.1]
85	25.72	30.4	(24.9, 27.4]
86	31.79	34.3	(30.7, 34.2]

87	27.16	31.8	(24.9, 27.4]
88	30.73	33.6	(30.7, 34.2]
89	37.22	43	(37.1, 40]
90	39.03	47.1	(37.1, 40]
91	34.36	40.3	(34.2, 37.1]
92	31.38	37.7	(30.7, 34.2]
93	24.35	29.3	(22.1, 24.9]
94	30.69	36	(30.7, 34.2]
95	37.82	53.9	(37.1, 40]
96	37.92	47.4	(37.1, 40]
97	27.79	30.9	(27.4, 30.7]
98	37.75	45.2	(37.1, 40]
99	32.5	39.9	(30.7, 34.2]
100	29.47	36.2	(27.4, 30.7]
