

Fredagsopgaven

Indledende kode:

```
ods graphics/imagemap=on;

libname ssvs "/courses/d284cd65ba27fe300/Sommerskole/Data";

proc contents data=ssvs.c87c46;
run;

data fredag;
set ssvs.c87c46;
    if ldep2 =. then delete;
    if vare =. then delete;
    if ldist =. then delete;
    if wgt_cont =. then delete;
    if ltotalli =. then delete;
    if bind =. then delete;
    if cwagr =. then delete;
    if cargo =. then delete;
    if mch =. then delete;
    if portms =. then delete;
id+1;
run;

proc contents data=fredag;
run;

proc freq data=fredag;
    table ABBREVIA HS;
run;
```

1

a

Kode:

```
proc reg data=fredag;
model ldep2=vare ldist wgt_cont ltotalli bind cwagr cargo mch portms;
run;
```

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-1.84753	1.09200	-1.69	0.0911
vare		1	-2.67529	0.11167	-23.96	<.0001
LDIST	log of distance from foreign port to mai	1	0.43166	0.10734	4.02	<.0001
WGT_CONT	Containerization (percentage of weight c	1	1.04446	0.33656	3.10	0.0020
LTOTALLI	total liner imports (in log)	1	-0.09023	0.02398	-3.76	0.0002
BIND	Price-fixing carrier agreements dummy (1	1	1.20929	0.16696	7.24	<.0001
CWAGR	Cooperative carrier agreements dummy (1	1	0.64211	0.13604	4.72	<.0001
CARGO	Cargo reservation dummy	1	-0.32015	0.14608	-2.19	0.0287
MCH	Cargo handling services index	1	-0.00455	0.21876	-0.02	0.9834
PORTMS	Mandatory port services index	1	0.54902	0.27150	2.02	0.0435

Der er en positiv sammenhæng mellem afstanden til havnen og prisen. Beta-estimatet for ldist er på 0,43166. Dette giver intuitivt god mening, da jo længere varene transporteres jo dyrere bliver det. Dette er pga. flere reparationer, mere olie, mandskab mv.

b

Kode:

```
data fredag;
set ssvs.c87c46;
ldist2=ldist**2;
run;
```

```
proc reg data=fredag;
model ldep2=vare ldist ldist2 wgt_cont ltotalli bind cwagr cargo mch portms;
run;
```

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-14.87571	3.96579	-3.75	0.0002
vare		1	-2.64996	0.11105	-23.86	<.0001
LDIST	log of distance from foreign port to mai	1	3.99986	1.05025	3.81	0.0002
ldist2		1	-0.22885	0.06701	-3.42	0.0007
WGT_CONT	Containerization (percentage of weight c	1	0.63032	0.35530	1.77	0.0765
LTOTALLI	total liner imports (in log)	1	-0.09964	0.02395	-4.16	<.0001
BIND	Price-fixing carrier agreements dummy (1	1	1.00943	0.17570	5.75	<.0001
CWAGR	Cooperative carrier agreements dummy (1	1	0.75790	0.13918	5.45	<.0001
CARGO	Cargo reservation dummy	1	-0.34036	0.14507	-2.35	0.0193
MCH	Cargo handling services index	1	-0.02240	0.21713	-0.10	0.9179
PORTMS	Mandatory port services index	1	0.45677	0.27075	1.69	0.0921

Jeg opretter et interaktionsled, $ldist^2$. Det kan udledes, at ldist-estimatet bliver positivt større, mens $ldist^2$ giver et negativt estimat på -0,22885. Det kan tolkes som, at marginalomkostningen på kort sigt er stigende,

mens det efter en hvis afstand/speciel havn er billigere. Fx. kan marginalomkostningerne være voksende indtil 1000 km sejllads, hvorefter det bliver billigere. Dette kan også kaldes stordriftsfordele.

2

a+b

Kode:

```
proc reg data=fredag plots=none;
model ldep2 = vare ldist ldist2 wgt_cont ltotalli bind cwagr cargo mch portms/selection=b ;
output out=fredag1;
run;
```

```
proc reg data=fredag plots=none;
model ldep2 = vare ldist ldist2 wgt_cont ltotalli bind cwagr cargo mch portms;
test mch=0;
run;
```

Test 1 Results for Dependent Variable LDEP2				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.02016	0.01	0.9179
Denominator	681	1.89512		

Ved først og køre en backwards selection test. Her finder jeg en enkelt insignifikant variable, mch, som derfor droppes. Dette tester jeg bagefter med en test. Denne bekræfter, at mch-variablen skal droppes.

c

Variablen *bind* beskriver en dummy for samarbejdspriser med andre virksomheder. Dette er en mark-up-pricing tillagt virksomhedens egne marginalomkostninger. Dette sikrer, at alle ydelser udbydes til markedspris. Hvis dummyen er til stede, så vil prisen stige. *cwagr* er en dummy for samarbejde mellem virksomheder. Hvis samarbejdet er til stede, så vil prisen stige. Dette ses, da begge parameter-estimer er positive. Dette giver tegn op karteldannelse. Dette tyder på, at hvis virksomhederne arbejder sammen på pris og andre parameter, så vil prisen stige. Da begge parametre er signifikante, vil det have en effekt.

3

a

```
%let sejeskibe = vare ldist ldist2 wgt_cont ltotalli bind cwagr cargo portms;
```

```
proc reg data=fredag1 plots=none;
id name;
model ldep2 = &sejeskibe./partial influence;
output out=fredag2 Rstudent=t covratio=c h=h cookd=d;
run;
```

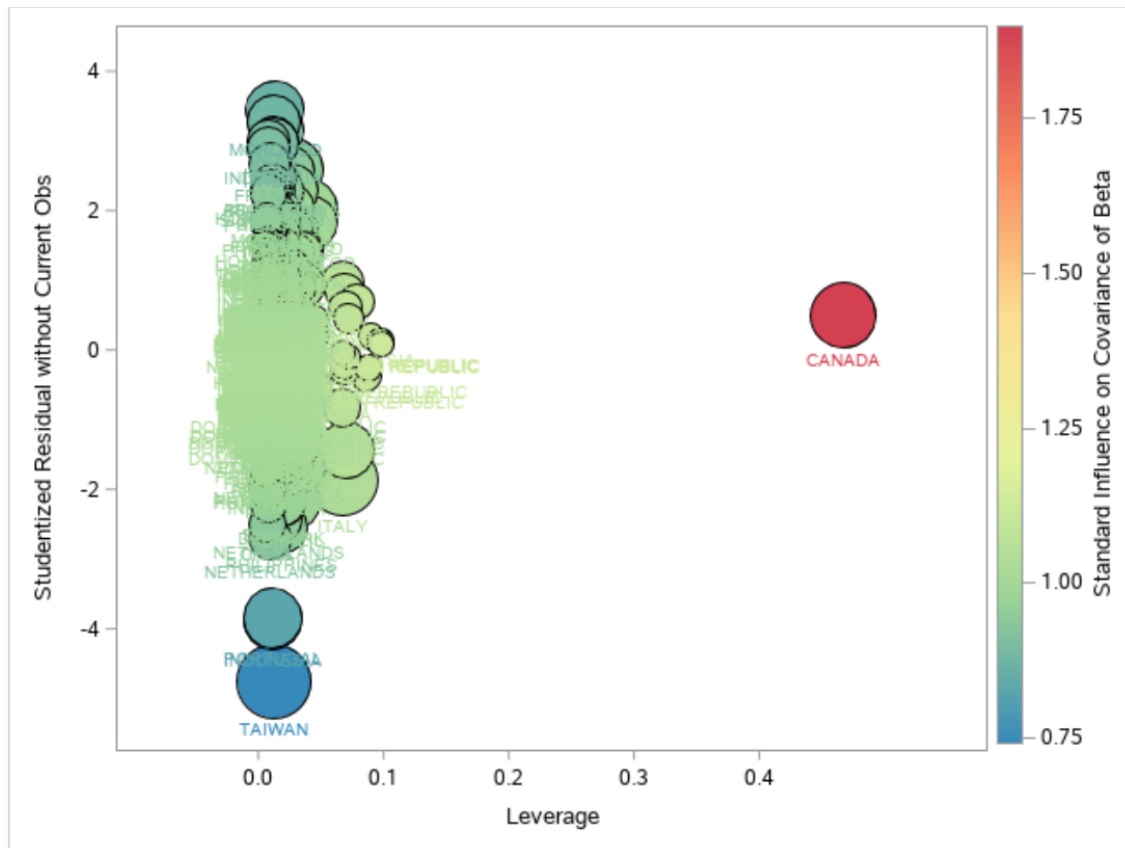
```

proc sort data=fredag2 out=fredag3;
by t;
run;

proc print data=fredag3;
var name t h c d;
where abs (d)>0.01;
run;

proc sgplot data=fredag2;
bubble x=h y=t size=d / datalabel=name datalabelpos=bottom colorresponse=c
colormodel=(CX3288BD CX99D594 CXE6F598 CXFEE08B CXFC8D59 CXD53E4F);;
run;

```



Det kan udledes, Canada har en stor indflydelse på grafen. Dette kan udledes, fordi den har stor leverage og en minimal afvigelse fra modellen. (Residualen er lille). Dette er derfor et godt punkt. Modsat er fx Taiwan, Portugal og Indonesien nogle mindre gode punkter. De har lav leverage, og de afviger mest fra modellen. De kan derfor senere fjernes.

b

```

proc print data=fredag3;
var name t h c d;
where abs (t)>3;
run;

```

```
%let n = 692;
%let k = 10;
%let t = 4.75008;
%let p = 0.05;
```

```
data fredag2;
p1=probt( &t.,(&n.-1-&k.));
p2=2*(1-p1);
p3=1-(1-p2)**&n.;
tgraense=tinv( 1- ( 1- ( 1 - &p.)*(1/&n.) )/2, &n.-1-&k.);
tbonf=tinv( 1- &p./2/&n., &n.-1-&k.);
run;
proc print;
run;
```

Obs	p1	p2	p3	tgraense	tbonf
1	1.00000	.000002481	.001715249	3.98709	3.99328

Det kan udledes ud fra p_3 -værdien, at sandsynligheden for at have en outlier såsom Taiwan fønevnt er $0,0017 \approx 0,17\%$. Altså er sandsynligheden meget lille.

c

Kode:

```
proc sort data=fredag2 out=fredag4;
by c;
run;
```

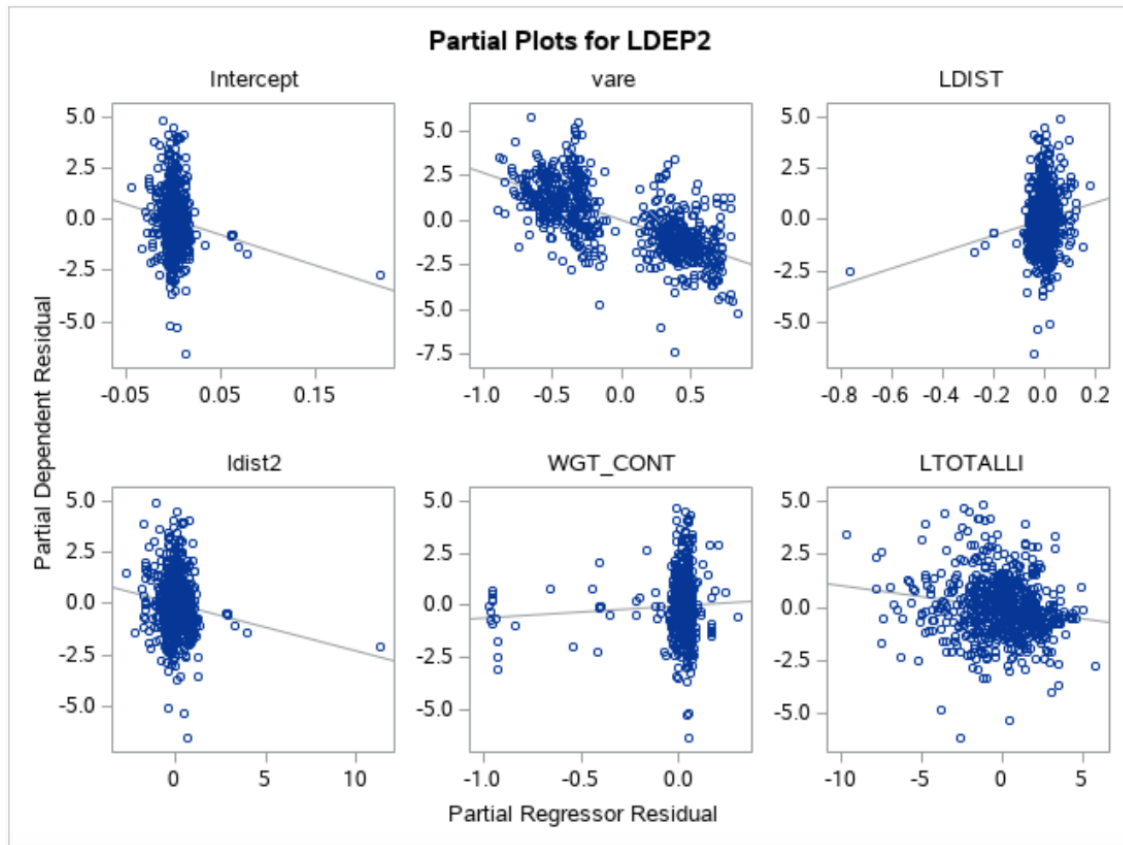
```
proc print data=fredag4;
var name t h c d;
where abs (c)<0.9;
run;
```

Obs	NAME	t	h	c	d
1	TAIWAN	-4.75008	0.012849	0.74204	0.028469
2	INDONESIA	-3.87692	0.011286	0.82507	0.016812
3	PORTUGAL	-3.84500	0.011462	0.82814	0.016803
4	MOROCCO	3.45161	0.013266	0.86468	0.015765
5	INDIA	3.28072	0.012177	0.87819	0.013080
6	SPAIN	3.14475	0.013857	0.89088	0.013718
7	ITALY	2.99686	0.008018	0.89736	0.007176

Det kan ses, at Taiwan, Indonesien og Portugal observationerne igen fremgår. Altså er de dårlige parametre på alle punkter. Derfor kan de fjernes for at gøre modellen bedre, men så kan man ende ud i en situation, hvor man ikke har særlig mange observationer tilbage, fordi man har fjernet det hele.

d

```
proc reg data=fredag1 plots=all;
id name;
model ldep2 = &sejeskibe./partial;
run;
```

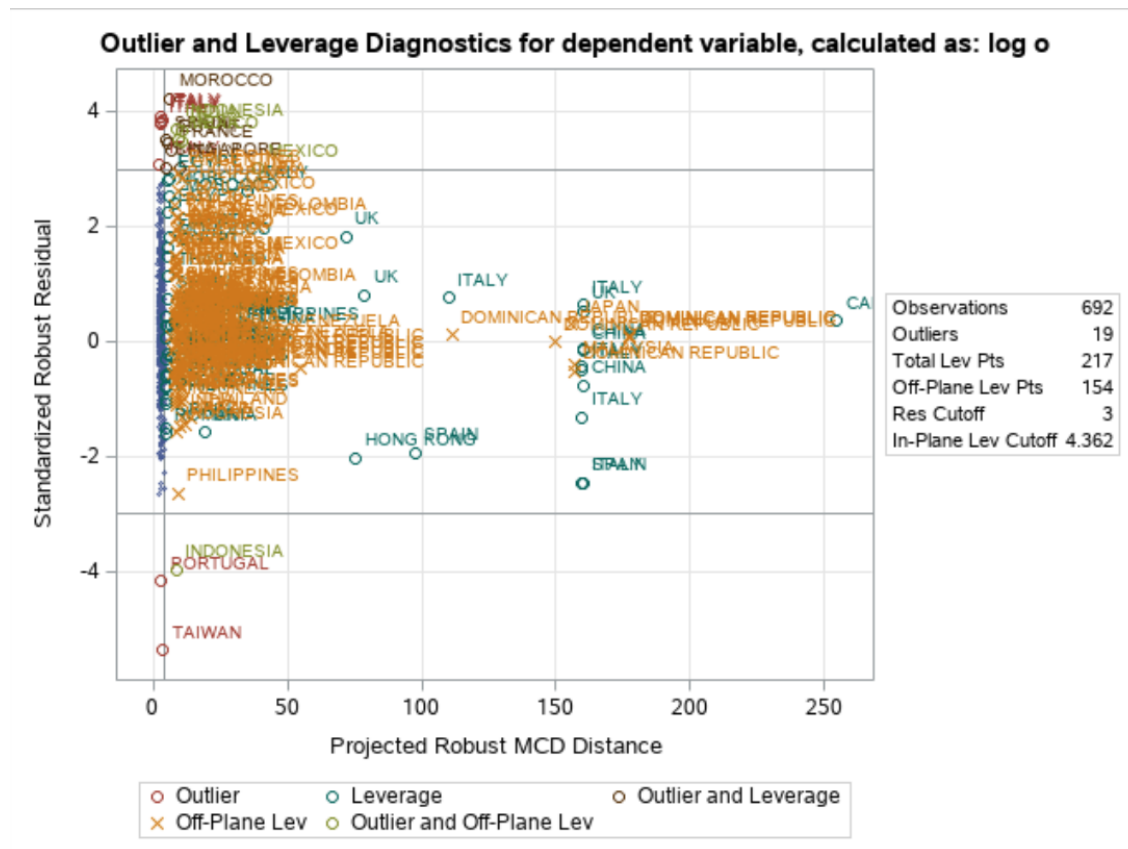


De ligner lidt de forrige plots. Det ses, at $ldist^2$ er negativ og $ldist$ er positiv. Begge ting stemmer overens tidligere observationer.

4

a

```
%let alleskibe = vare ldist ldist2 wgt_cont ltotalli bind cwagr cargo mch portms;
proc robustreg data=fredag1 method=lts plots=all;
model ldep2 = &alleskibe./diagnostics(all);
id name;
output out=fredag6 outlier=yafvig leverage=xafvig;
run;
```



Det kan udledes, at der er 19 outliers. Leverage cutoff ved 4,362. Det observeres, at der er 217 leverage punkter. Igen kan det udledes, at punkterne Indonesia, Portugal og Taiwan er outliers, ligesom de har været dårlige punkter gennem hele opgaven.

b

```
proc reg data=fredag6 plots=none;
model ldep2 = &alleskibe./influence;
id name;
where (yafvig=0 and xafvig=0);
```

```
proc reg data=fredag6 plots=none;
model ldep2 = &alleskibe./influence;
id name;
where yafvig=0;
```

```
proc reg data=fredag6 plots=none;
model ldep2 = &alleskibe./influence;
id name;
where xafvig=0;
run;
```

Root MSE	1.14906	R-Square	0.6662
Dependent Mean	0.42000	Adj R-Sq	0.6597
Coeff Var	273.58499		

Figure 1: X og Y

Root MSE	1.24844	R-Square	0.6510
Dependent Mean	0.46286	Adj R-Sq	0.6442
Coeff Var	269.72219		

Figure 2: X

Root MSE	1.20339	R-Square	0.6250
Dependent Mean	0.36819	Adj R-Sq	0.6193
Coeff Var	326.83757		

Figure 3: Y

Det kan udledes, at fjernes både leverage punkter og outliere, så bliver $R^2_{y=x=0}$ den højeste. Sætter man kun $y = 0$, så vil $R^2_{y=0}$ -værdien falde til 0,6250, hvilket er mindre end 0,6682, når man sætter begge variable lig nul. Hvis man kun sætter $x = 0$, så vil $R^2_{x=0}$ -værdien være 0,6510.

c

Det kan ses ud fra *partial plots* lavet i opgave 3.d), at der er enkelte outliere. Dette stemmer overens med analysen i 4.b), hvor det ses, at ved at fjerne outlierne, så opnåes den model med den højeste R^2 -værdi. Da jeg ikke kan aflæse mine partial plots gør dette min analyse en smule hvis ikke meget vag.

d

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-203.57492	44.78974	-4.55	<.0001
vare		1	-2.58794	0.11254	-23.00	<.0001
LDIST	log of distance from foreign port to mai	1	36.38993	8.35490	4.36	<.0001
ldist2		1	-2.03091	0.47386	-4.29	<.0001
WGT_CONT	Containerization (percentage of weight c	1	43.05401	27.98677	1.54	0.1246
LTOTALLI	total liner imports (in log)	1	-0.04507	0.02798	-1.61	0.1079
BIND	Price-fixing carrier agreements dummy (1	1	1.76838	0.24495	7.22	<.0001
CWAGR	Cooperative carrier agreements dummy (1	0	0	.	.	.
CARGO	Cargo reservation dummy	1	-0.23206	0.17965	-1.29	0.1971
MCH	Cargo handling services index	1	-0.51634	0.30420	-1.70	0.0903
PORTMS	Mandatory port services index	1	-0.16536	0.36364	-0.45	0.6495

I forhold til tidligere, så kommer der færre variable, der er signifikante. Nu droppes ogs *wgtcont*, *ltotalli*, *cargoogportms*. Dette er *og leverage-punkter er droppet*. Både estimatorne og varians har ændret sig, hvilket giver mening, da leverage punkter er fjernet.

5

a+b+c

```
proc glm data=;
class vare ;
model ldep2 = vare
vare*ldist vare*bind vare*cwagr vare*cargo/solution;
run;
quit;
```


Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	-15.81402619	B	4.64724108	-3.40	0.0007
vare 0	-4.64953510	B	8.99946337	-0.52	0.6056
vare 1	0.00000000	B	.	.	.
LDIST*vare 0	5.29172412		1.96053669	2.70	0.0071
LDIST*vare 1	4.01732759		1.28365442	3.13	0.0018
ldist2*vare 0	-0.29970365		0.12258876	-2.44	0.0147
ldist2*vare 1	-0.25472163		0.08408208	-3.03	0.0025
WGT_CONT*vare 0	-0.04409026		0.53699271	-0.08	0.9346
WGT_CONT*vare 1	1.22364771		0.47242333	2.59	0.0098
LTOTALLI*vare 0	-0.09520023		0.03116696	-3.05	0.0023
LTOTALLI*vare 1	-0.10780563		0.03728637	-2.89	0.0040
BIND*vare 0	1.06988577		0.22803416	4.69	<.0001
BIND*vare 1	0.53221450		0.25700683	2.07	0.0388
CWAGR*vare 0	0.64631658		0.24653210	2.62	0.0089
CWAGR*vare 1	0.91896090		0.17822501	5.16	<.0001
CARGO*vare 0	-0.14050853		0.21430197	-0.66	0.5123
CARGO*vare 1	-0.52640967		0.17891025	-2.94	0.0034
PORTMS*vare 0	1.62988113		0.43556796	3.74	0.0002
PORTMS*vare 1	-0.41096572		0.34011211	-1.21	0.2273

På de to forskellige varer får jeg forskellig påvirkning for at shipping-prisen øges. Hvis $vare = 0$, så er bind meget højere end hvis $vare = 1$. Begge er signifikante. Altså, på vare 0 er markuppen på prisen højere end ved vare 1 for variabelen bind. Modsat er det på $cwagr = 0 \wedge cwagr = 1$. Her er der større påvirkning på vare 1. Igen er begge signifikante.

Jeg ser ikke nogen steder på de signifikante variable, at parameterestimaterne har de forkerte fortegn. Så skal det være, at $ldist^2$ er negativt, hvilket tyder på samme fortolkning for distance som tidligere nævnt. Altså, at over en hvis længde, så bliver det billigere at transportere varen. Dette kan være pga. de høje omkostninger ved korte sejlladser, super tankere taget i betragtning.

Konklusionen om karteldannelse ændres ikke, men det kan dog tydes, at der forskel af påvirkning på prisen alt efter varen, der fragtes, og hvor fra den fragtes.

6

a+b

```
proc logistic data=fredag1 plots=(effectplot);
model bind = ldep2 vare ldist ldist2 wgt_cont ltotalli cwagr cargo portms/influence;
id name;
run;
```

```
proc logistic data=fredag1 plots=(effectplot);
model cwagr = ldep2 vare ldist ldist2 wgt_cont ltotalli bind cargo portms/influence;
id name;
run;
```

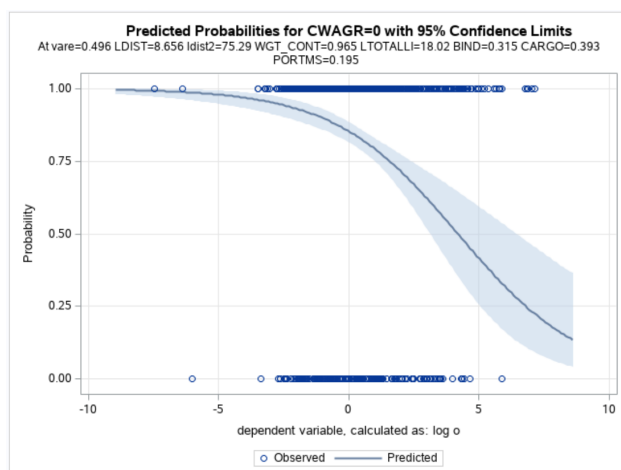


Figure 4: Cwagr

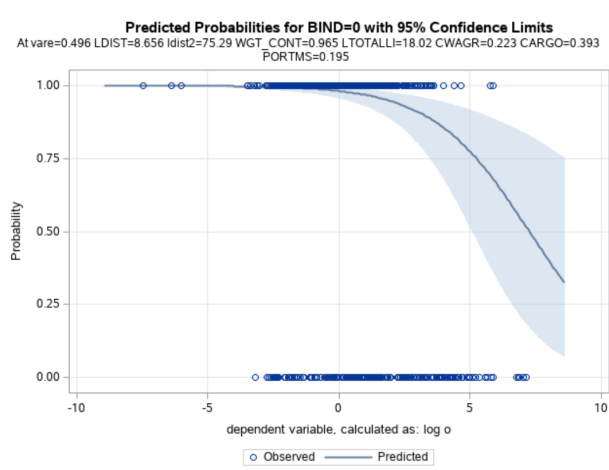


Figure 5: Bind

Herved ses det, at sandsynligheden fra observationerne ift. $bind = 0 \wedge cwagr = 0$, er faldende i begge parameter. Hermed er sandsynligheden faldende ved prisen, når man ikke er i kartel. Altså falder prisen, når man ikke er i kartel. Igen har jeg svært ved at aflæse mine kurver, men det kan ses, der er nogle punkter i begge plots, der stikker en smule ud, hvorfor de må trække i fordelingen. Med her hentydes til de punkter, der fuldt synlige med en bæå cirkel rundt om noget hvidt.

7

a+b

```
data fredag7;
set fredag1;
aftale=0;
if bind = 1 and cwagr=0 then aftale=1;
if bind = 0 and cwagr=1 then aftale=2;
if bind = 1 and cwagr=1 then aftale=3;
run;
proc logistic data=fredag7 plots=all;
model aftale = ldep2 vare ldist ldist2 wgt_cont ltotalli cargo mch portms;
run;

proc logistic data=fredag7 plots=all;
model aftale = ldep2 vare ldist ldist2 wgt_cont ltotalli cargo mch portms
/link=logit;
run;
```

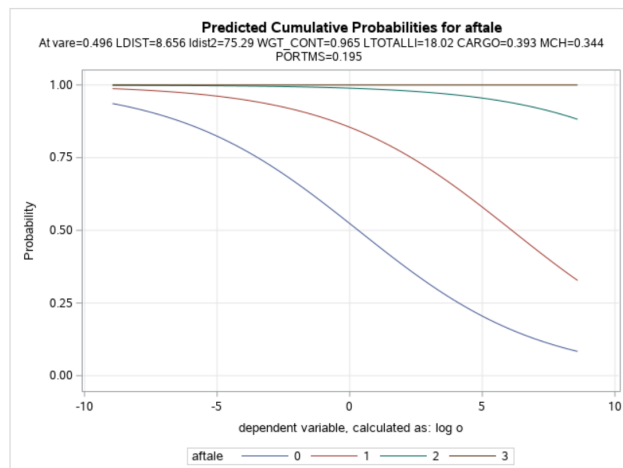


Figure 6: Ikke-ordnet

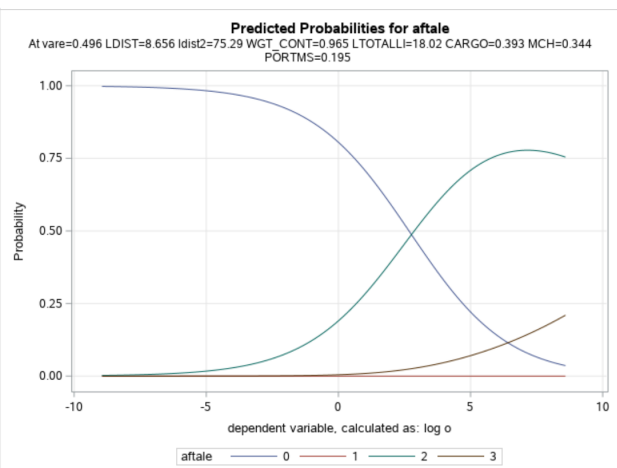


Figure 7: Ordnet

For den ordnede fordeling(til højre) kan det ses, at