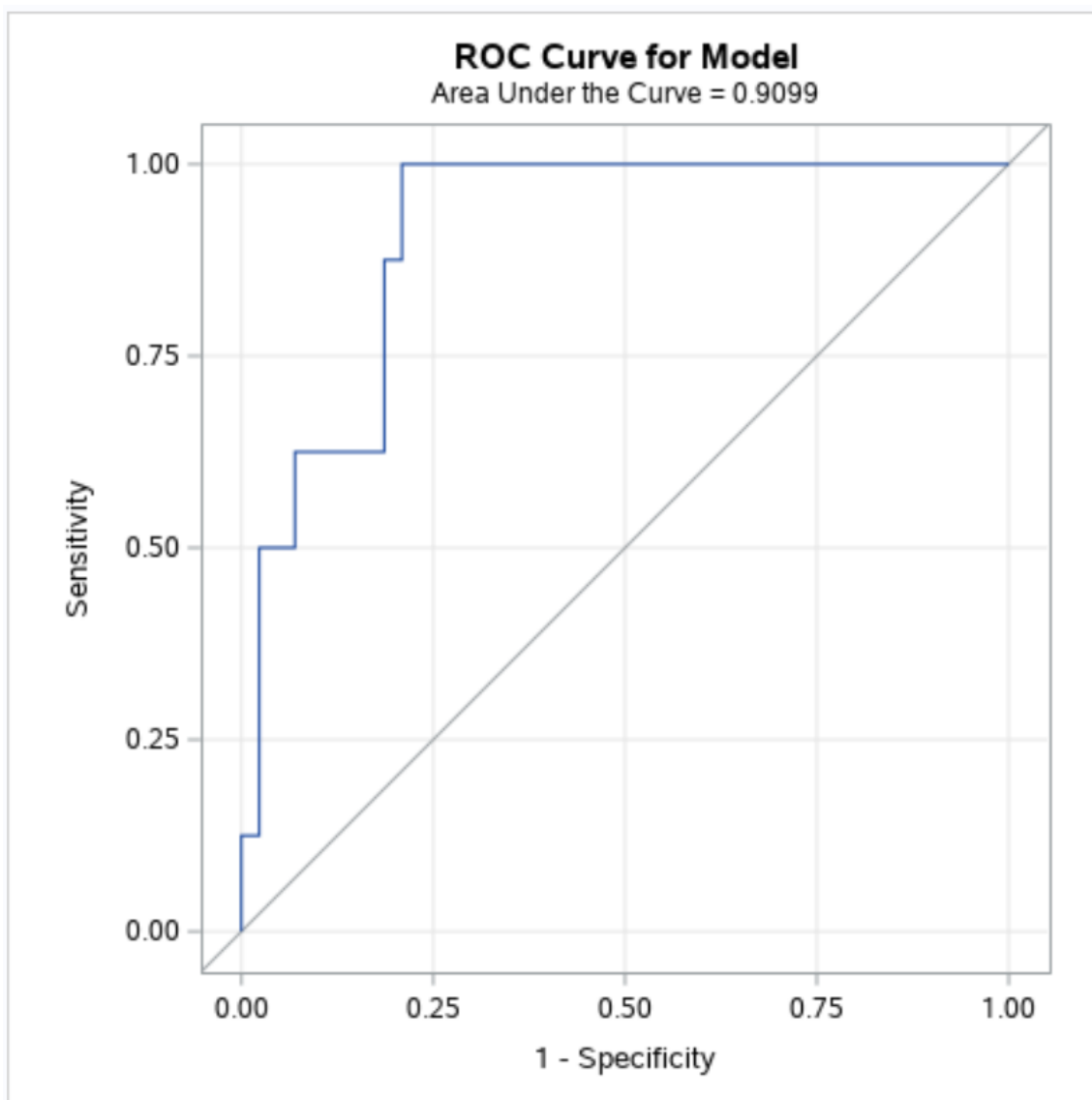


## Torsdagsopgaven

a

Kode:

```
ods graphics/imagemap=on;  
  
libname ssvs "/courses/d284cd65ba27fe300/Sommerskole/Data";  
  
proc contents data=ssvs.usa_2022;  
run;  
  
proc logistic data=ssvs.usa_2022 plots=all;  
model cat=rate_black ;  
run;
```



Det kan udledes, at vi gør brug af variabelen "black". Altså måles sandsynligheden for at være "black". Area under the Curve er givet ved 0,9099, hvorved vi kan konkludere, det er en god model, vi arbejder med.

**b**

Kode:

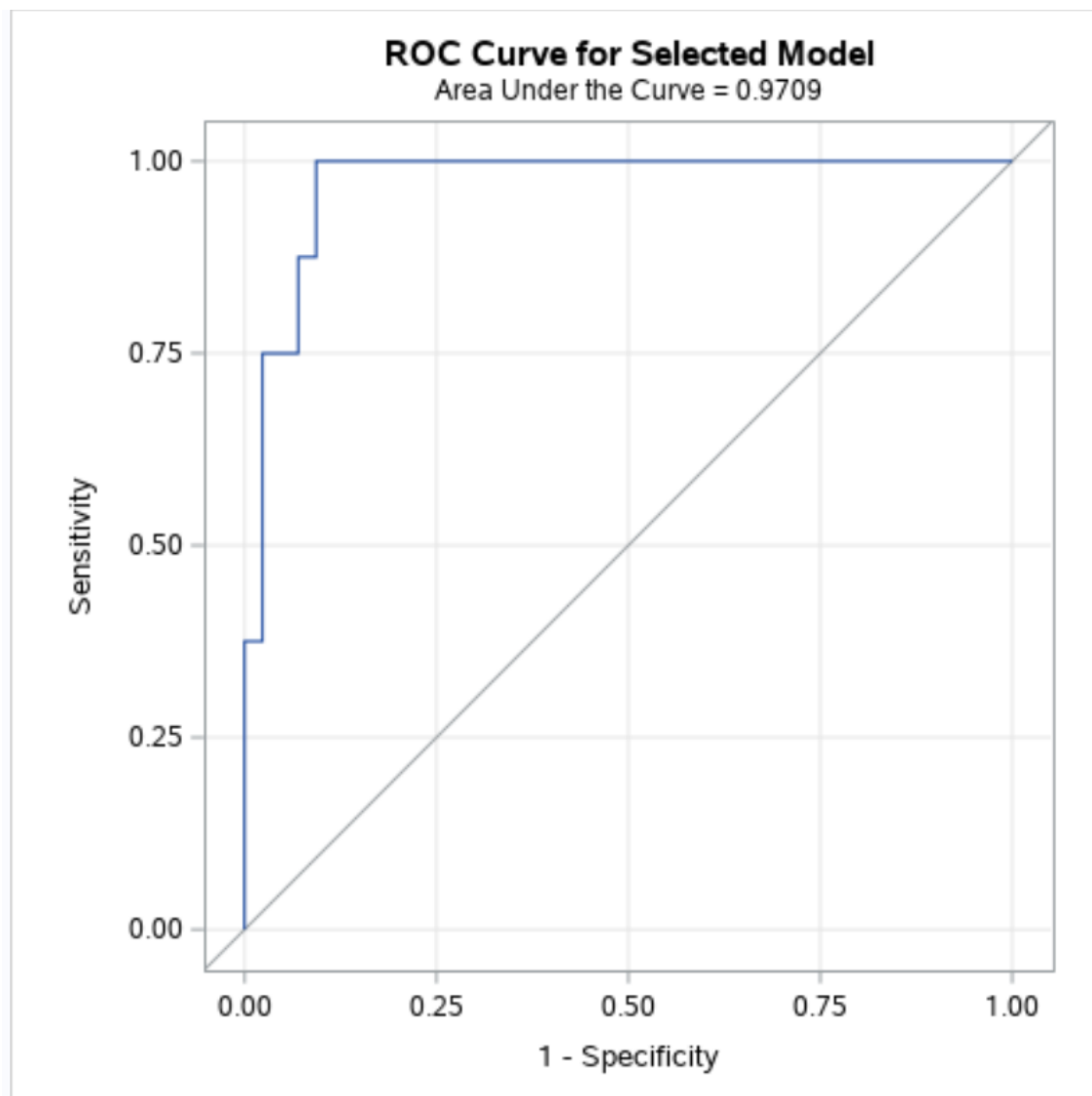
```
%let var1 = percent_biden trump_2016 rate_black income_2018 rate_income_growth unemployment;

proc logistic data=ssvs.usa_2022 plots=all;
model cat=&var1.;
run;

proc logistic data=ssvs.usa_2022 plots=all;
model cat=&var1./selection=b;
run;

%let var2 = rate_black trump_2016;
%put _user_;

proc logistic data=ssvs.usa_2022 plots=all;
model cat = &var2. / influence;
effectplot fit(x=rate_black) /at (trump_2016= 0 trump_2016= min midrange max );
id state;
score out = usa_1 clm;
run;
```

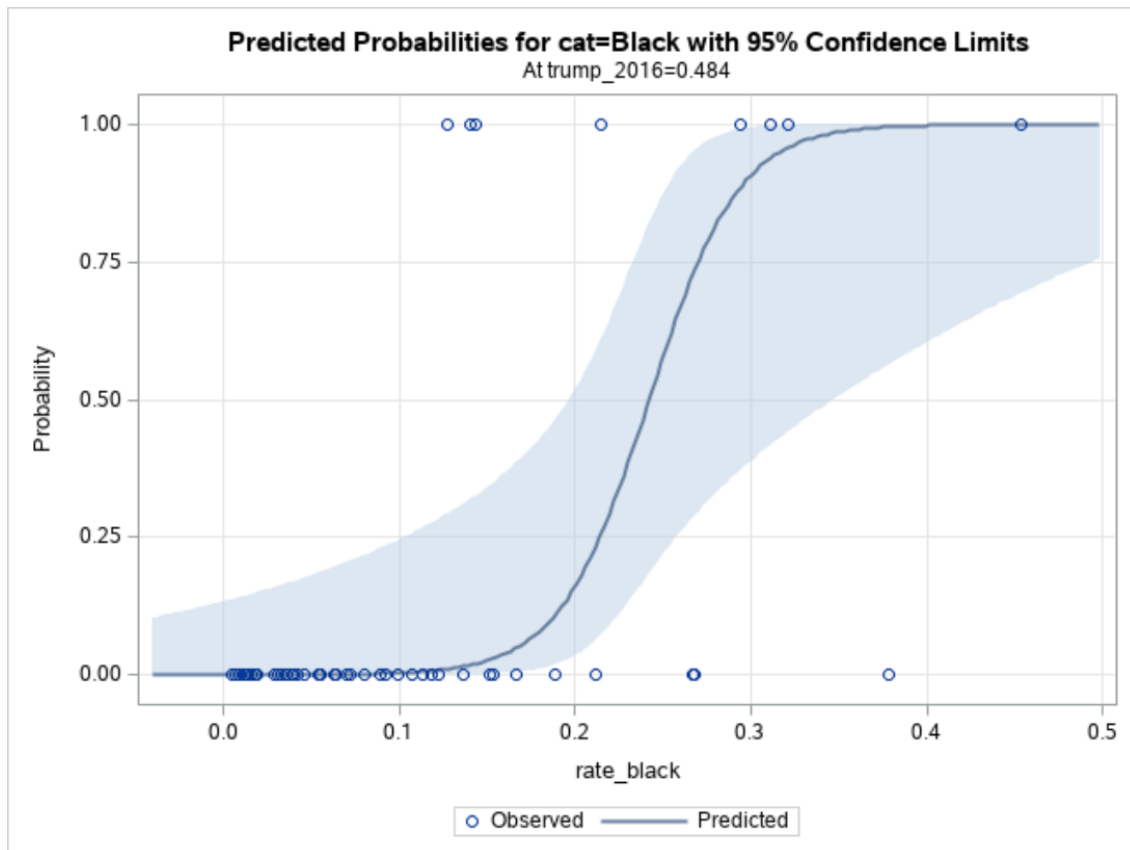


Det kan udledes, at arealet under grafen er givet ved 0.9709.

| Analysis of Maximum Likelihood Estimates |    |          |                |                 |            |
|--|----|----------|----------------|-----------------|------------|
| Parameter                                | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept                                | 1  | 9.6550   | 5.5296         | 3.0487          | 0.0808     |
| rate_black                               | 1  | 39.4735  | 16.8117        | 5.5130          | 0.0189     |
| trump_2016                               | 1  | -39.7132 | 18.9183        | 4.4066          | 0.0358     |

Når rateblack variablen stiger med 1 procent, så stiger sandsynligheden for at cat er sort med 39%. Hvis man modsat stemte på Trump i 2016, så falder sandsynligheden for at staten har flest sorte med 39%. Begge

værdier er signifikante, da jeg vha. backwards selection har fjernet de insignifikante variable.



**c**

Kode:

```
*c*;

data Score;
set score_1;
label
cat = "Faktisk andel"
i_cat="Forventet andel"
p_black="Forventet Andel dr bte sorte";
Forsk = abs(rate_black/100-p_black);
Korrekt = "Ja"; if i_cat = cat then Korrekt="Nej";
run;

proc sort data=score;
by Korrekt forskel;
run;

proc print data=score label;
var state Korrekt p_Black cat i_cat Forskel;
run;
```

```
proc print data=score label;
var state Korrekt p_Black cat i_cat Forskel;
where cat ne i_cat;
run;
```

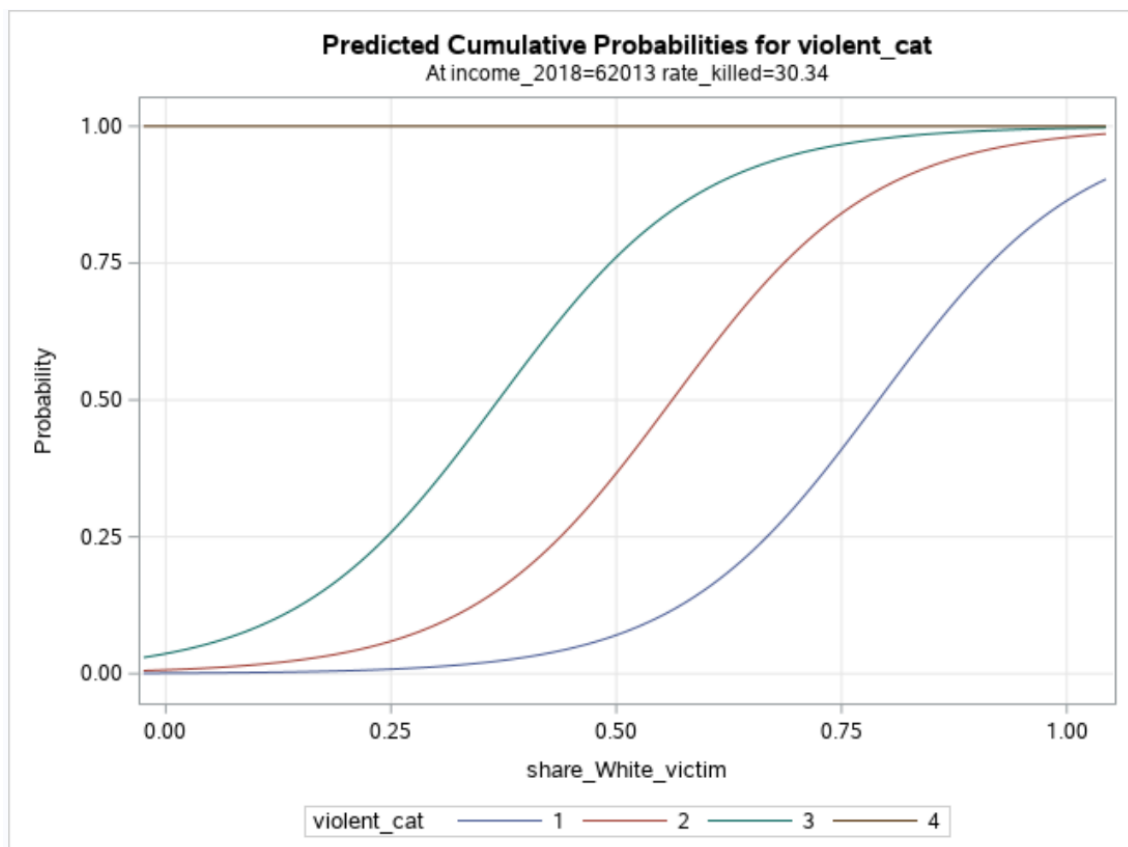
| Obs | State       | Korrekt | Forventet Andel dræbte sorte | Faktisk andel | Forventet andel | Forskel |
|-----|-------------|---------|------------------------------|---------------|-----------------|---------|
| 1   | New Jersey  | Ja      | 0.16881                      | Black         | White           | 0.16754 |
| 2   | Louisiana   | Ja      | 0.32655                      | Black         | White           | 0.32334 |
| 3   | Illinois    | Ja      | 0.44597                      | Black         | White           | 0.44457 |
| 4   | Mississippi | Ja      | 0.83161                      | White         | Black           | 0.82782 |

Det kan udledes, at modellen gætter forkert 4 gange. Derfor er modellen ikke helt præcis. Dette kan omhandle mængden af sorte, der bor i staterne. Specielt Louisiana og Mississippi er tæt befolket med sorte/african-american mennesker. Herved kan det være svært at gætte ud fra, hvor mange, der stemmer på trump og andelen af sorte.

## d1

Kode:

```
proc logistic data=ssvs.usa_2022 plots=all;
model violent_cat = share_Black_victim share_White_victim rate_black percent_biden trump_2020
/selection=b;
run;
```



| Analysis of Maximum Likelihood Estimates |   |    |          |                |                 |            |
|--|---|----|----------|----------------|-----------------|------------|
| Parameter                                |   | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept                                | 1 | 1  | -10.0751 | 3.2765         | 9.4556          | 0.0021     |
| Intercept                                | 2 | 1  | -8.0452  | 3.1399         | 6.5650          | 0.0104     |
| Intercept                                | 3 | 1  | -6.3382  | 3.0745         | 4.2499          | 0.0393     |
| share_White_victim                       |   | 1  | 8.8558   | 2.2382         | 15.6554         | <.0001     |
| income_2018                              |   | 1  | 0.000094 | 0.000036       | 6.9134          | 0.0086     |
| rate_killed                              |   | 1  | -0.0917  | 0.0258         | 12.5960         | 0.0004     |

Man måler sandsynligheden ift. til at være i gruppe 4, hvor der er mest kriminalitet. Det ses, at jo flere, der dræbes i staten, jo større sandsynlighed er der for at være i gruppen med mest kriminalitet. Hvis der modsat er et højt niveau af hvide, der dræbes, i staten, så mindskes sandsynligheden for at være i den mindst kriminelle gruppering. Det er svært at konkludere noget på variabelen income, da måleenheden ikke er givet.

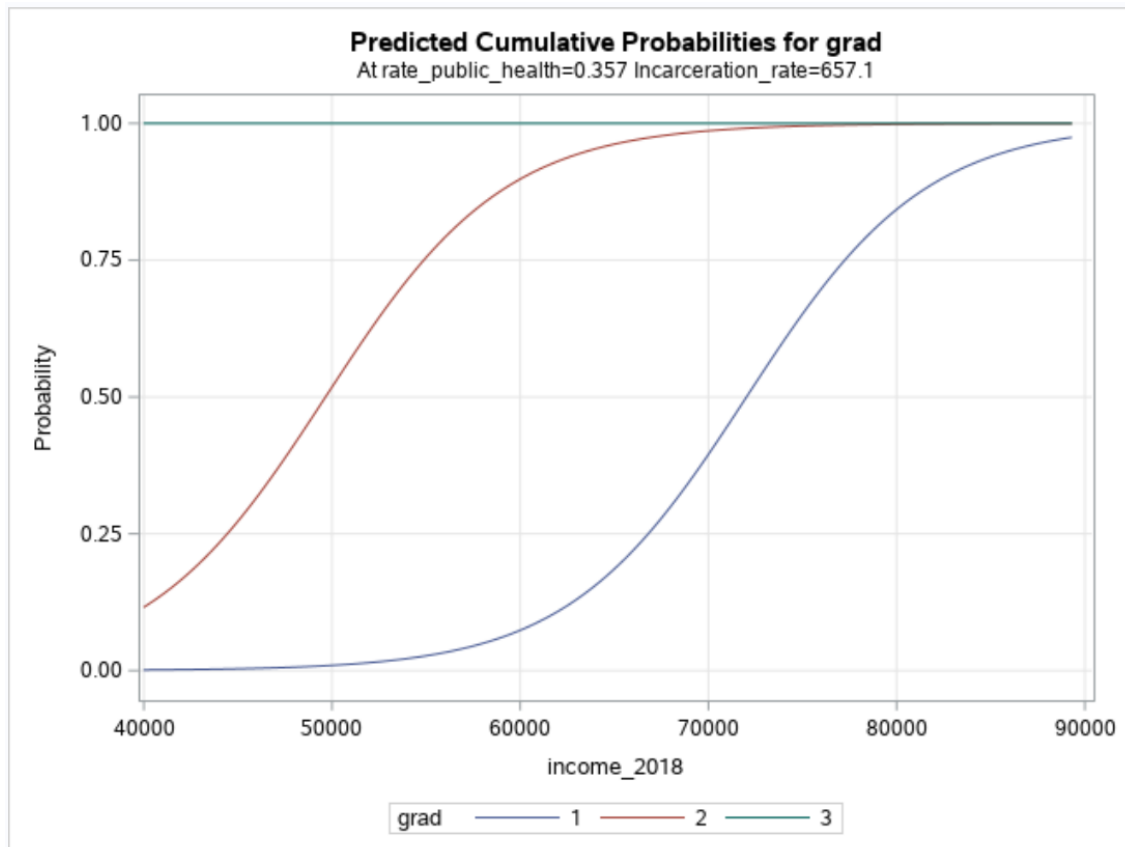
Det kan udledes, at score-testen. Ssh for at parameter er den samme i gennem den de fire modeller. Man har regerensen, næst mest mellem krome. HVad er ssh for at parameteren stiger, når man går fra 1 til 2. Dog giver score-testen os en værdi på 0,9227 something, hvilket er meget større end 0,05, hvorved den går igennem hele modellen.

## d2

Kode:

```
proc logistic data=ssvs.usa_2022 plots=all;
model grad = share_Black_victim share_White_victim rate_black
income_2018 rate.income.growth unemployment_rate
rate_public_health rate_non_insured_health
rate_private_health Incarceration_rate rate_killed/selection=b;
run;
```

| Analysis of Maximum Likelihood Estimates |   |    |          |                |                 |            |
|--|---|----|----------|----------------|-----------------|------------|
| Parameter                                |   | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept                                | 1 | 1  | -22.9232 | 6.4842         | 12.4979         | 0.0004     |
| Intercept                                | 2 | 1  | -18.2151 | 5.8856         | 9.5783          | 0.0020     |
| income_2018                              |   | 1  | 0.000210 | 0.000058       | 13.1904         | 0.0003     |
| rate_public_health                       |   | 1  | 33.2989  | 9.0508         | 13.5358         | 0.0002     |
| Incarceration_rate                       |   | 1  | -0.00626 | 0.00225        | 7.7497          | 0.0054     |

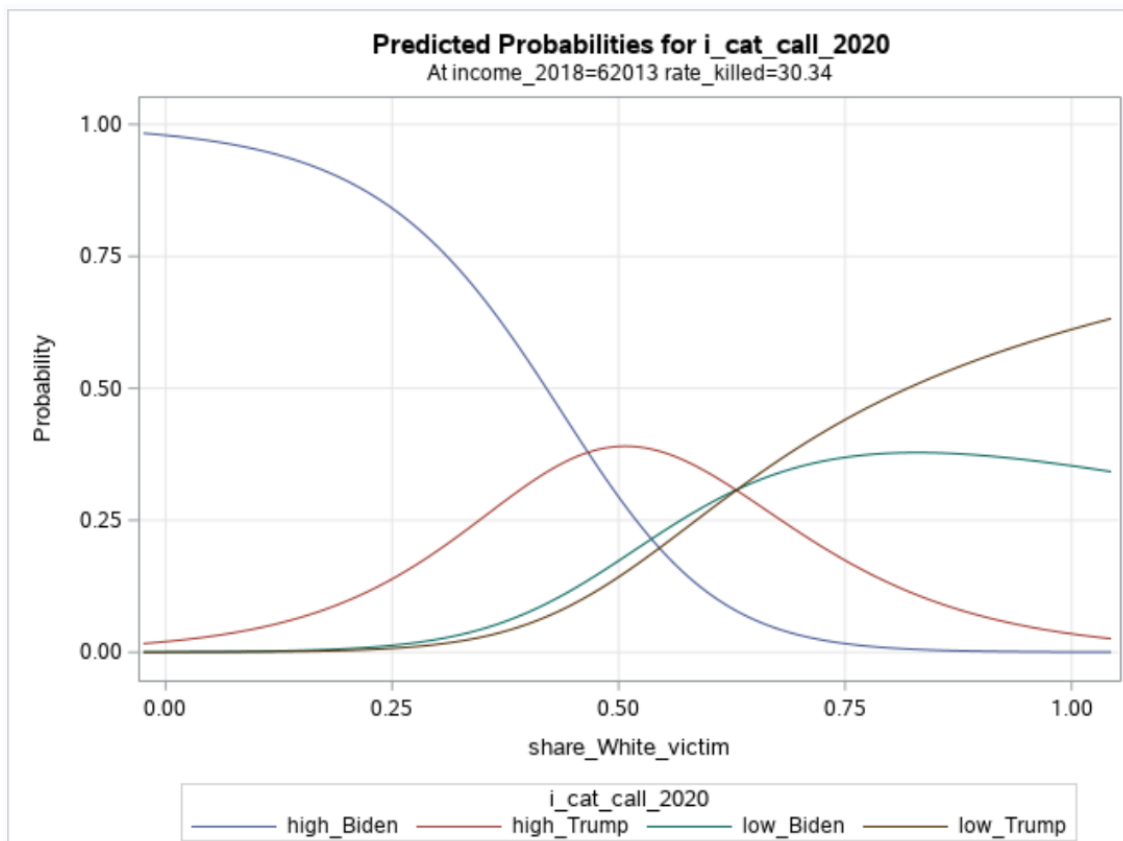


Det kan ses, at de signifikante variable er income 2018, rate public health og incarceration rate. Ud fra disse ses det, at højere andel af indsatte giver større sandsynlighed for at være mod abort. Jo bedre det offentlige sundhedssystem er, jo mere for abort er man, mens income igen er svært at konkludere på. Dette er pga måleenheden ikke er sikker. Ifølge parallelitetsantagelsen som måler om estimator berører hele vejen igennem. Dette testes på et 5% signifikans-niveau. Denne forkastes pga score-testen viser 0.9836. Herved afvises det.

e

Kode:

```
proc logistic data=ssvs.usa_2022 plots=(effectplot);
model i_cat_call_2020 = share_Black_victim share_White_victim rate_black
income_2018 rate_income_growth unemployment_rate
rate_public_health rate_non_insured_health
rate_private_health Incarceration_rate rate_killed/selection=b link=glogit;
id state;
run;
```



| Analysis of Maximum Likelihood Estimates |                 |    |          |                |                 |            |
|--|-----------------|----|----------|----------------|-----------------|------------|
| Parameter                                | i_cat_call_2020 | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept                                | high_Biden      | 1  | 3.6564   | 6.1587         | 0.3525          | 0.5527     |
| Intercept                                | high_Trump      | 1  | 6.4588   | 5.4498         | 1.4046          | 0.2360     |
| Intercept                                | low_Biden       | 1  | -9.3360  | 6.2880         | 2.2044          | 0.1376     |
| share_White_victim                       | high_Biden      | 1  | -16.0798 | 5.0960         | 9.9565          | 0.0016     |
| share_White_victim                       | high_Trump      | 1  | -7.7424  | 3.7475         | 4.2683          | 0.0388     |
| share_White_victim                       | low_Biden       | 1  | -1.4881  | 3.9182         | 0.1442          | 0.7041     |
| income_2018                              | high_Biden      | 1  | 0.000048 | 0.000082       | 0.3439          | 0.5576     |
| income_2018                              | high_Trump      | 1  | -0.00005 | 0.000074       | 0.5280          | 0.4675     |
| income_2018                              | low_Biden       | 1  | 0.000207 | 0.000086       | 5.8104          | 0.0159     |
| rate_killed                              | high_Biden      | 1  | 0.0703   | 0.0494         | 2.0213          | 0.1551     |
| rate_killed                              | high_Trump      | 1  | 0.0583   | 0.0463         | 1.5814          | 0.2086     |
| rate_killed                              | low_Biden       | 1  | -0.0854  | 0.0587         | 2.1137          | 0.1460     |

Her ses effekterne af high trump, high Biden og low Biden ift sandsynligheden for andelen af hvide ofre i staterne. Vi ser, at hvis der er en lille del af hvide ofre, så er der høj sandsynlighed for at være high Biden. Jo større andelen af hvide, der dør i staten, jo større er sandsynligheden for low Trump.



f

Kode:

```
proc genmod data=usa plots=all;
model total=percent_biden trump_2016 income_2018 rate_black
unemployment_rate rate_private_health rate_income_growth
rate_public_health rate_non_insured_health Incarceration_rate
/dist=p offset=log_population;
run;
```

| Analysis Of Maximum Likelihood Parameter Estimates |    |          |                |                            |         |                 |            |
|--|----|----------|----------------|----------------------------|---------|-----------------|------------|
| Parameter  | DF | Estimate | Standard Error | Wald 95% Confidence Limits |         | Wald Chi-Square | Pr > ChiSq |
| Intercept  | 1  | -7.7944  | 1.1675         | -10.0827                   | -5.5061 | 44.57           | <.0001     |
| percent_biden                                      | 1  | -0.0589  | 0.0062         | -0.0710                    | -0.0468 | 90.87           | <.0001     |
| trump_2016   | 1  | -6.5675  | 0.6251         | -7.7927                    | -5.3424 | 110.39          | <.0001     |
| income_2018  | 1  | 0.0000   | 0.0000         | 0.0000                     | 0.0000  | 13.90           | 0.0002     |
| rate_black   | 1  | -1.7174  | 0.1754         | -2.0611                    | -1.3737 | 95.92           | <.0001     |
| unemployment_rate                                  | 1  | -0.0747  | 0.0232         | -0.1201                    | -0.0292 | 10.36           | 0.0013     |
| rate_private_health                                | 1  | 0.1794   | 0.9187         | -1.6213                    | 1.9801  | 0.04            | 0.8452     |
| rate_income_growth                                 | 1  | -0.4023  | 0.3445         | -1.0775                    | 0.2729  | 1.36            | 0.2429     |
| rate_public_health                                 | 1  | 3.5309   | 1.0094         | 1.5526                     | 5.5092  | 12.24           | 0.0005     |
| rate_non_insured_heal                              | 1  | 2.9671   | 1.3248         | 0.3704                     | 5.5637  | 5.02            | 0.0251     |
| Incarceration_rate                                 | 1  | 0.0022   | 0.0001         | 0.0019                     | 0.0024  | 311.20          | <.0001     |
| Scale  | 0  | 1.0000   | 0.0000         | 1.0000                     | 1.0000  |                 |            |

Ovenstående er de givne estimator ud fra regressionen. Variablen total er antal dræbte af politiet mellem 2013 og 2022. Det giver ikke rigtig mening, at ved faldende arbejdsløshed stiger antallet af dræbte. Hvilket lyder mærkværdigt. Samme ses med black, jo flere sorte, jo færre politidrab. Dette giver ikke rigtig mening, da politiet i USA nærmest jager sorte mennesker. Det kan også ses, bedre offentlig sundhed giver flere politidrab. Denne fordeling stemmer ikke overens med virkeligheden.