Stat 3302 Report

Group H: Sam Lefebvre.22, Nina Wei.1241, Maggie Miller.10302, Kevin Song.1649

4/20/2023

Introduction

In the article Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes by John Paul Wilson and Nicholas O. Rule, an experiment is done to see if people's perception of trustworthiness in criminals could be a good predictor for extreme criminal-sentencing outcomes. In order to make the study valid, they wanted to do this with actual murderers, instead of just running an experiment with perceived guilt. The goal was to see if people giving a trust rating to each of these people was relevant. In the first study, people were shown murderers and asked to rate them on trustworthiness on a scale from 1 to 8. In the second study, the same people were shown people who were exonerated from death row. It was hypothesized that the results would show that these ratings would have a correlation to extreme criminal sentencing outcomes. In the first study, photos of inmates were collected from the Florida Department Corrections. These are people who were convicted of murder and are now waiting on death row. Females were excluded as well as males who were not black or white, which left them with 371 men. They then selected a control group of similar looking men who had life in prison, resulting in a total of 742 images that were then divided into 7 smaller sets. After, 208 workers from Amazon Mechanical Turk who were all American were then asked to rate these photos on how trustworthy they assumed they were on a scale from 1 to 8 with 1 being not trustworthy and 8 being extremely trustworthy. Other samples were also taken to rate the subjects in the photos based on Afrocentricity, attractiveness and facial maturity. Some other factors that were also taken into account were fWHR (facial width to height ratio), presence of glasses, and presence of tattoos. In the second study, photos were collected from the innocence project of people who have been exonerated from their crimes. From those, 37 men were selected whose crimes would've either landed them on death row or life in prison. The same group of workers was then brought in to rate these photos using the same traits as study one, while also factoring in how many years the person spent in prison. For the results for study 1, all of the ratings were averaged out in order to get a group judgment for each rating for each face. It was found that subjects who were sentenced to death looked less trustworthy than those who were sentenced to life in prison. Black subjects were also rated more trustworthy than white targets, but race and sentence did not interact. A logistic regression model was then built to see the extent that facial trustworthiness would predict the sentencing outcomes as well as with the other traits that were measured. It was determined that targets who were seen as less trustworthy had a higher likelihood to be sentenced to death than people who were seen as more trustworthy. Trustworthiness did seem to be a good predictor, but Afrocentricity, fWHR and glasses were also significant. In the second study, another logistic regression model was built. Again, it was found that trustworthiness did a good job predicting sentencing outcomes. In this model, none of the covariates predicted death sentencing. The final results after both of the studies concluded were that facial appearance really does affect criminal sentencing independent of actual guilt, and also the severity of the sentence. Overall, this points out a problem in the criminal-justice system that they need to be made aware of, or bias on the grounds of facial features will continue to have an impact on sentencing. The goals of our report are to replicate the data that was produced by the experiment in both study 1 and study 2. Each of these studies has two models within them that account for all of the characteristics that have been previously listed. We will be doing this by creating two binary logistic regression models for each study, and then displaying the data using the summary function. The results of both of the studies that are predicting sentencing outcomes will be shown, which should match those that were given in the article.

Methods

There was one primary methodology to analyze the relationship between facial trustworthiness and actual criminal sentences across two studies. Both studies involved the gathering of faces of convicted murderers who had been sentenced to either death or life in prison.

Study 1

In Study 1, the facial trustworthiness of every death-row inmate in Florida was analyzed in order to investigate the relationship between facial trustworthiness and sentencing among real criminals serving life sentences or awaiting execution. Stimulus selection was utilized as 376 white and black males were selected, excluding female inmates and inmates of other races. This was done in order to keep sex and race uniform. A control group of inmates sentenced to life imprisonment was selected to compare against death row inmates.

Next, in order to collect data on the facial traits of the inmates, defendants were divided into seven sets, each set containing equal life sentences and death row sentences, as well as racial splits. They obtained trustworthiness ratings by asking 208 American workers from Amazon Mechanical Turk to rate their trustworthiness from 1 to 8. Ratings of other facial traits were collected, indluding Afrocentricity, attractiveness, facial maturity, glasses, and tattoos. These trustworthiness ratings were compared to a facial width-to-height ratio (fWHR) rating (aggression rating) using ImageJ, and there was a small negative correlation between them, suggesting an independent relationship.

Two logistic regression models were used in order to determine the relationship of facial trustworthiness and predicted sentencing outcomes. The first regressed sentence outcome (0 = life, 1 = death) onto trustworthiness. The second added more covariates in addition to trustworthiness. The covariates were Afrocentricity, attractiveness, facial maturity, fWHR, presence of glasses, and presence of tattoos.

To reproduce the two models, we will use the `glm()` R function with the `family=binomial` parameter to create binary logistic regression models that regress sentence outcome onto the corresponding covariates for each model.

Study 2

Study 2 tested the relationship between perceived trustworthiness and criminal sentencing among innocent individuals. Stimulus selection was utilized to select from a group of 107 people. They recorded information such as the sentence received, length of time served, crime committed, year of conviction, and state the conviction occurred. They eliminated people who lived in a state without a death sentence, and selected people whose crimes would have made them eligible for the death sentence. This resulted in a total of 37 people, 20 of whom sentenced to life and 17 to death.

American workers from the Amazon Mechanical Turk were again asked to rate each of the 37 people for trustworthiness, Afrocentricity, attractiveness, and facial maturity. These were the same scales as Study 1. Presence of glasses and tattoos were coded for but fWHR (aggression) was unable to be recorded because targets were not uniformly facing the camera.

Two logistic regression models were used in order to determine if facial trustworthiness predicted sentencing outcomes for innocent individuals. The first regressed sentence outcome (0 = life, 1 = death) onto trustworthiness. The second added more covariates in addition to trustworthiness. The covariates were Afrocentricity, attractiveness, facial maturity, the presence of glasses, and the number of years each person served in prison.

To reproduce the two models, we will use the `glm()` R function with the `family=binomial` parameter to create binary logistic regression models that regress sentence outcome onto the corresponding covariates for each model.

Results

First, we import the two datasets needed for the two studies. `study1` is the dataset used for Study 1, and `study2` is the dataset used for Study 2.

```r
study1 = read.csv("Study1Data.csv")
study2 = read.csv("Study2Data.csv")
```

Study 1

Study 1 has two binary logistic regression models, Model 1 and Model 2. We were able to reproduce Model 1 and Model 2 in terms of coefficients and odds ratios as presented in Table 1.

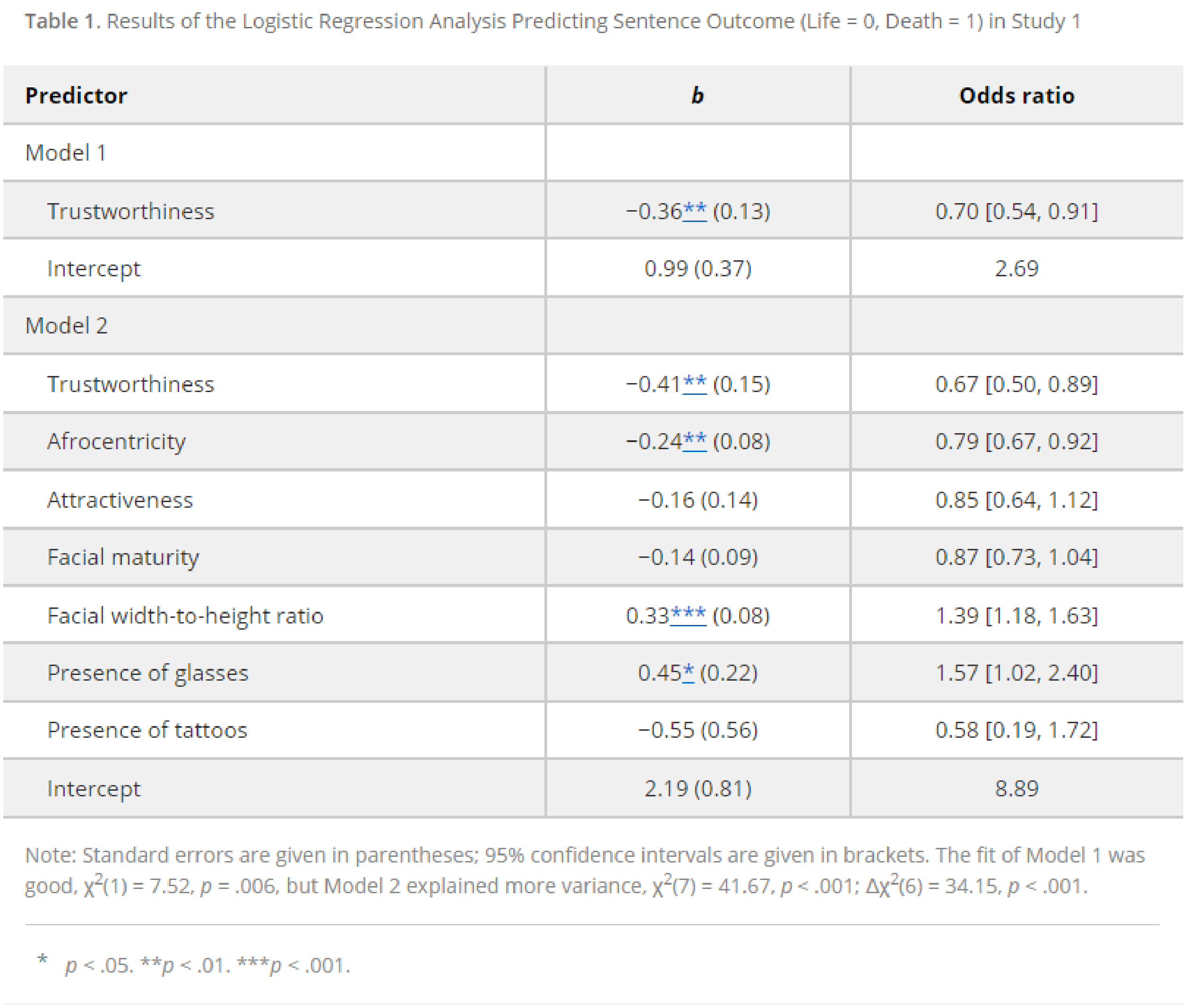


Table 1. Results of the Logistic Regression Analysis Predicting Sentence Outcome (Life = 0, Death = 1) in Study 1

Predictor	*b*	Odds ratio
Model 1		
Trustworthiness	−0.36** (0.13)	0.70 [0.54, 0.91]
Intercept	0.99 (0.37)	2.69
Model 2		
Trustworthiness	−0.41** (0.15)	0.67 [0.50, 0.89]
Afrocentricity	−0.24** (0.08)	0.79 [0.67, 0.92]
Attractiveness	−0.16 (0.14)	0.85 [0.64, 1.12]
Facial maturity	−0.14 (0.09)	0.87 [0.73, 1.04]
Facial width-to-height ratio	0.33*** (0.08)	1.39 [1.18, 1.63]
Presence of glasses	0.45* (0.22)	1.57 [1.02, 2.40]
Presence of tattoos	−0.55 (0.56)	0.58 [0.19, 1.72]
Intercept	2.19 (0.81)	8.89

Note: Standard errors are given in parentheses; 95% confidence intervals are given in brackets. The fit of Model 1 was good, $\chi^2(1) = 7.52$, $p = .006$, but Model 2 explained more variance, $\chi^2(7) = 41.67$, $p < .001$; $\Delta\chi^2(6) = 34.15$, $p < .001$.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Model 1 The results of Model 1 show that convicted murderers perceived as less trustworthy were more likely to be sentenced to death since the coefficient of `Trustworthiness` is negative.

```
s1mod1 = glm(sent ~ trust, family=binomial, data=study1)
summary(s1mod1)
```

```
##
## Call:
## glm(formula = sent ~ trust, family = binomial, data = study1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.37084  -1.17343   0.06317   1.16488   1.40920
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9860     0.3708   2.659  0.00783 **
## trust        -0.3542     0.1306  -2.712  0.00669 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1028.6  on 741  degrees of freedom
## Residual deviance: 1021.1  on 740  degrees of freedom
## AIC: 1025.1
##
## Number of Fisher Scoring iterations: 4
```

```
b = coef(s1mod1)
names(b) = c("Intercept", "Trustworthiness")
odds.ratio = exp(b)
s1mod1.df = as.data.frame(cbind(b, odds.ratio))
kable(s1mod1.df)
```

	b	odds.ratio
Intercept	0.9859922	2.680470
Trustworthiness	-0.3542279	0.701715

Model 2 The results of Model 2 show that faces perceived as less trustworthy were more likely to be sentenced to death since the coefficient of `Trustworthiness` is negative. Afrocentricity was negatively associated with the death sentence, and fWHR and the presence of glasses were positively associated with the death sentence.

```
s1mod2 = glm(sent ~ trust + zAfro + attract + maturity + zfWHR + glasses + tattoos, family=binomial, da
summary(s1mod2)
```

```
##
## Call:
## glm(formula = sent ~ trust + zAfro + attract + maturity + zfWHR +
##     glasses + tattoos, family = binomial, data = study1)
##
## Deviance Residuals:
```

```
##      Min        1Q    Median        3Q       Max
## -1.72948  -1.12232   0.02201   1.11995   1.69473
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.17865    0.81294   2.680  0.00736 **
## trust       -0.40801    0.14727  -2.770  0.00560 **
## zAfro       -0.23819    0.07993  -2.980  0.00288 **
## attract     -0.16195    0.14203  -1.140  0.25416
## maturity    -0.13552    0.08842  -1.533  0.12537
## zfWHR        0.32643    0.08384   3.894 9.87e-05 ***
## glasses      0.44772    0.21875   2.047  0.04068 *
## tattoos     -0.54913    0.55832  -0.984  0.32535
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1028.63  on 741  degrees of freedom
## Residual deviance:  987.05  on 734  degrees of freedom
## AIC: 1003
##
## Number of Fisher Scoring iterations: 4
```

```
b = coef(s1mod2)
names(b) = c("Intercept", "Trustworthiness", "Afrocentricity", "Attractiveness", "Facial maturity", "Fa
odds.ratio = exp(b)
s1mod2.df = as.data.frame(cbind(b, odds.ratio))
kable(s1mod2.df)
```

	b	odds.ratio
Intercept	2.1786519	8.8343889
Trustworthiness	-0.4080089	0.6649730
Afrocentricity	-0.2381915	0.7880517
Attractiveness	-0.1619539	0.8504804
Facial maturity	-0.1355170	0.8732644
Facial width-to-height ratio	0.3264348	1.3860179
Presence of glasses	0.4477239	1.5647467
Presence of tattoos	-0.5491270	0.5774537

Study 2

Study 2 has two binary logistic regression models, Model 1 and Model 2. We were able to reproduce Model 1 and Model 2 in terms of coefficients and odds ratios as presented in Table 2. For Model 1, the summary table in the paper displayed the odds ratio of `Trustworthiness` as 0.56, which seems to be a typo, when it was actually 0.21 as stated in the preceding paragraph. We, however, did obtain an odds ratio of 0.21 for `Trustworthiness` in our reproduction of Model 1.

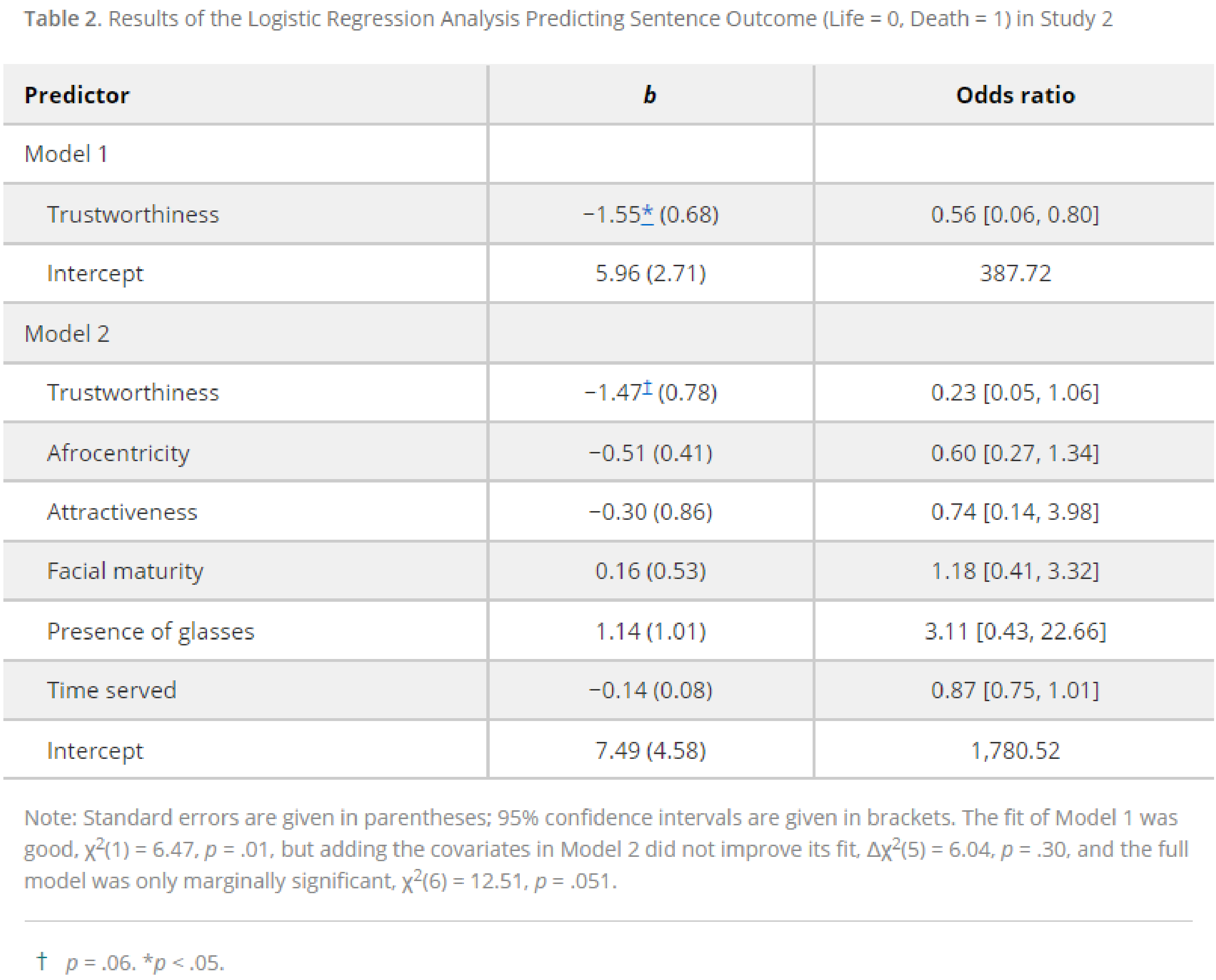


Table 2. Results of the Logistic Regression Analysis Predicting Sentence Outcome (Life = 0, Death = 1) in Study 2

Predictor	***b***	**Odds ratio**
Model 1		
Trustworthiness	−1.55* (0.68)	0.56 [0.06, 0.80]
Intercept	5.96 (2.71)	387.72
Model 2		
Trustworthiness	−1.47† (0.78)	0.23 [0.05, 1.06]
Afrocentricity	−0.51 (0.41)	0.60 [0.27, 1.34]
Attractiveness	−0.30 (0.86)	0.74 [0.14, 3.98]
Facial maturity	0.16 (0.53)	1.18 [0.41, 3.32]
Presence of glasses	1.14 (1.01)	3.11 [0.43, 22.66]
Time served	−0.14 (0.08)	0.87 [0.75, 1.01]
Intercept	7.49 (4.58)	1,780.52

Note: Standard errors are given in parentheses; 95% confidence intervals are given in brackets. The fit of Model 1 was good, $\chi^2(1) = 6.47$, $p = .01$, but adding the covariates in Model 2 did not improve its fit, $\Delta\chi^2(5) = 6.04$, $p = .30$, and the full model was only marginally significant, $\chi^2(6) = 12.51$, $p = .051$.

† $p = .06$. *$p < .05$.

Model 1 The results of Model 1 show that faces perceived as less trustworthy were more likely to be sentenced to death since the coefficient of `Trustworthiness` is negative.

```r
s2mod1 = glm(sent ~ trust, family=binomial, data=study2)
summary(s2mod1)
```

```
##
## Call:
## glm(formula = sent ~ trust, family = binomial, data = study2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2250  -0.8924  -0.6554   0.9818   1.6531
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.9708     2.7069   2.206   0.0274 *
## trust        -1.5513     0.6775  -2.290   0.0220 *
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 51.049  on 36  degrees of freedom
## Residual deviance: 44.556  on 35  degrees of freedom
## AIC: 48.556
##
## Number of Fisher Scoring iterations: 4
```

```
b = coef(s2mod1)
names(b) = c("Intercept", "Trustworthiness")
odds.ratio = exp(b)
s2mod1.df = as.data.frame(cbind(b, odds.ratio))
kable(s2mod1.df)
```

	b	odds.ratio
Intercept	5.970829	391.8303467
Trustworthiness	-1.551290	0.2119743

Model 2 In Model 2, adding the additional covariates did not improve the previous model (Model 1), and none of the covariates predicted the death sentence. However, `Trustworthiness` still remained the predictor that had an continuous effect on the death sentence.

```
s2mod2 = glm(sent ~ trust + zAfro + attract + maturity + glasses + served, family=binomial, data=study2)
summary(s2mod2)
```

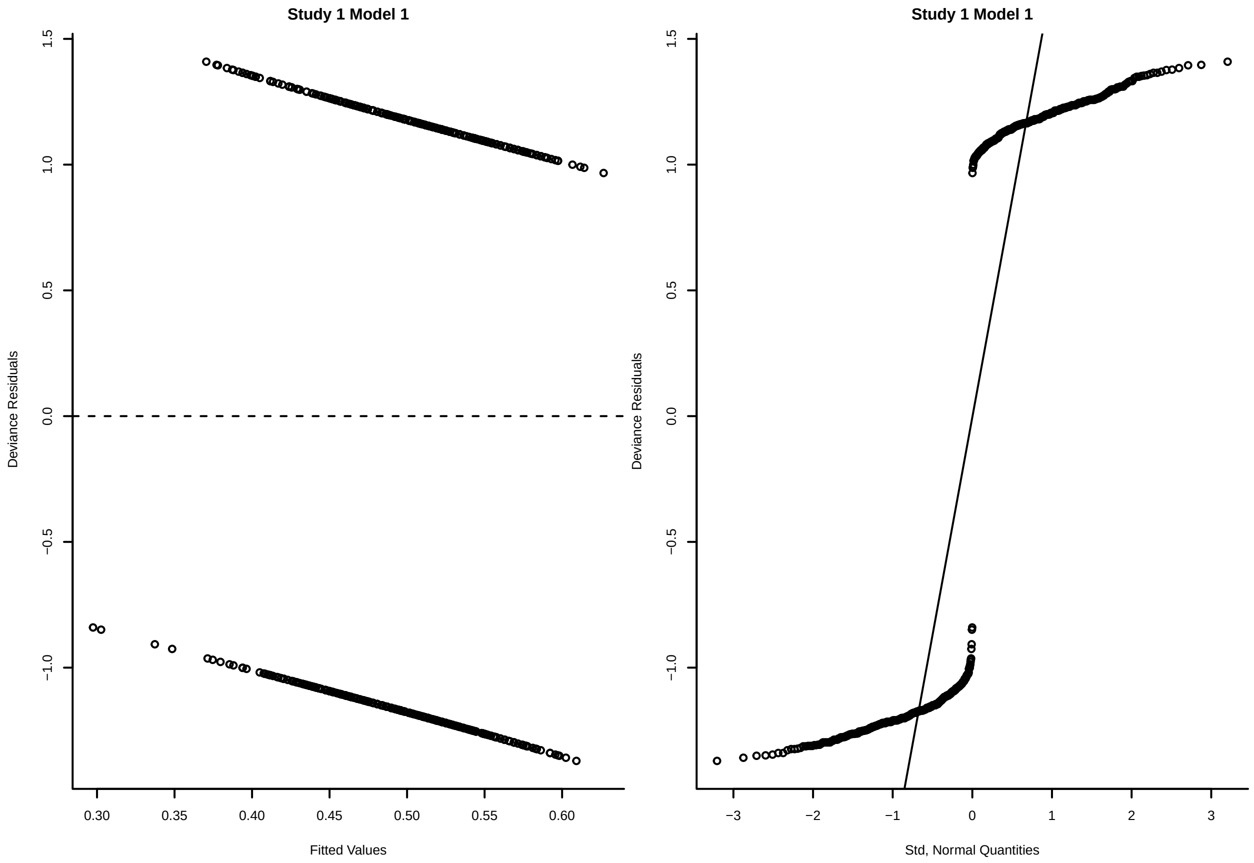
```
##
## Call:
## glm(formula = sent ~ trust + zAfro + attract + maturity + glasses +
##     served, family = binomial, data = study2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0928  -0.7751  -0.3472   0.8505   1.8089
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   7.4996     4.5693   1.641   0.1007
## trust        -1.4764     0.7826  -1.886   0.0592 .
## zAfro        -0.5139     0.4123  -1.246   0.2126
## attract      -0.3015     0.8584  -0.351   0.7254
## maturity      0.1606     0.5304   0.303   0.7620
## glasses       1.1375     1.0128   1.123   0.2614
## served       -0.1421     0.0777  -1.829   0.0674 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 51.049  on 36  degrees of freedom
## Residual deviance: 38.517  on 30  degrees of freedom
## AIC: 52.517
##
## Number of Fisher Scoring iterations: 4
```

```
b = coef(s2mod2)
names(b) = c("Intercept", "Trustworthiness", "Afrocentricity", "Attractiveness", "Facial maturity", "Pr
odds.ratio = exp(b)
s2mod2.df = as.data.frame(cbind(b, odds.ratio))
kable(s2mod2.df)
```

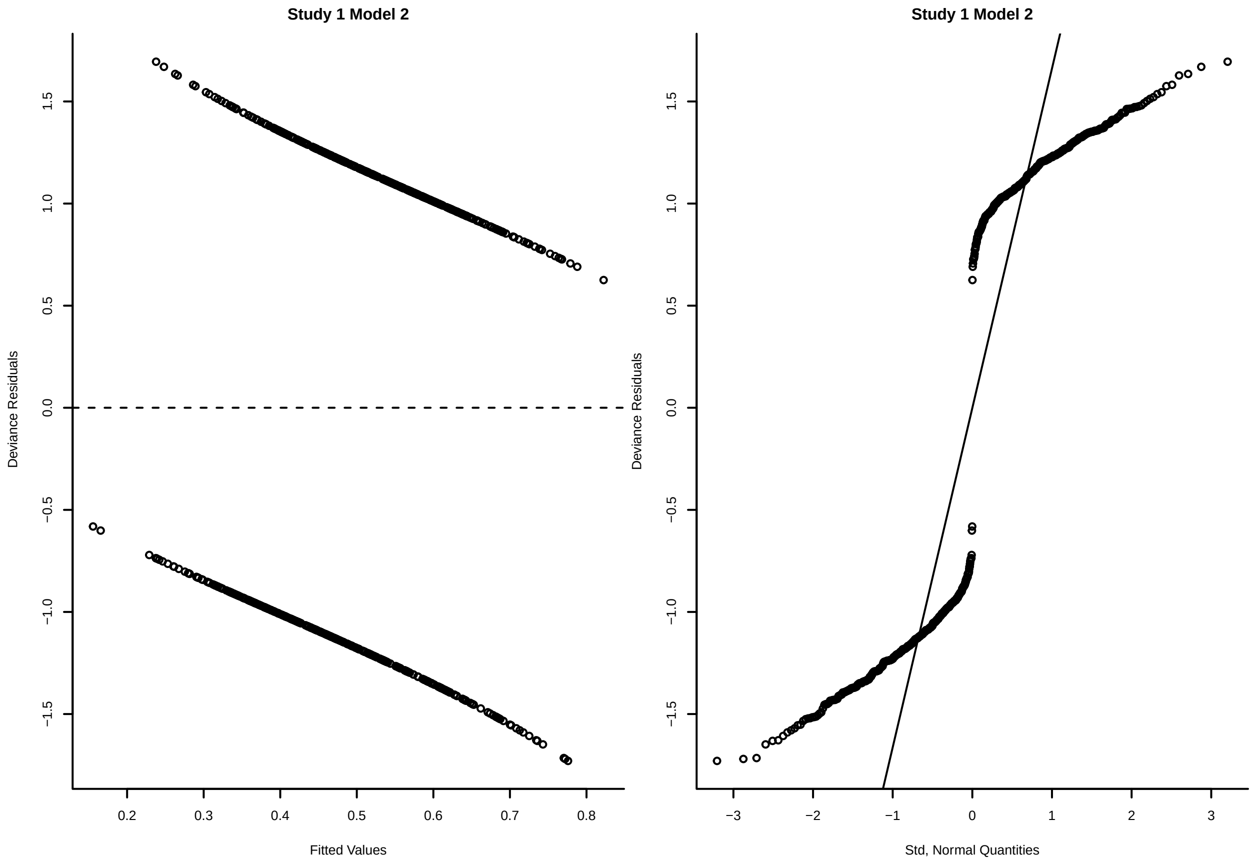
	b	odds.ratio
Intercept	7.4995939	1807.3083365
Trustworthiness	-1.4763530	0.2284694
Afrocentricity	-0.5138548	0.5981852
Attractiveness	-0.3015445	0.7396749
Facial maturity	0.1605948	1.1742091
Presence of glasses	1.1374907	3.1189321
Time served	-0.1421293	0.8675091

Commentary

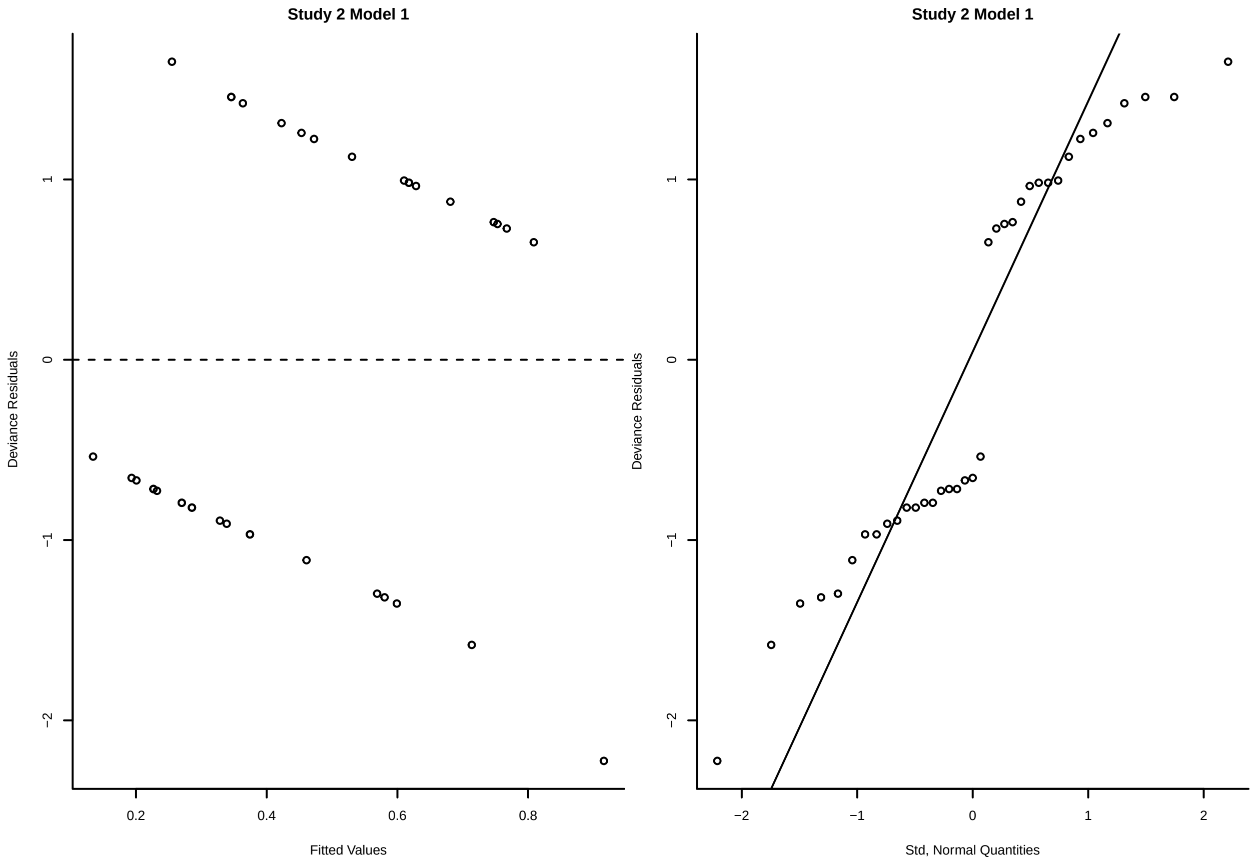
```
# Study 1 Plots
s1mod1.fits <- fitted(s1mod1)
s1mod1.dev.resids <- resid(s1mod1)
par(mfrow=c(1,2), cex =0.45, mar = c(4,4,2.3,0.2), bty ="L")
plot(s1mod1.fits, s1mod1.dev.resids, xlab = " Fitted Values ", ylab = " Deviance Residuals", main =  "S
abline(h=0, lty=2)
qqnorm(s1mod1.dev.resids, xlab = "Std, Normal Quantities", ylab = "Deviance Residuals", main = "Study 1
qqline(s1mod1.dev.resids)
```



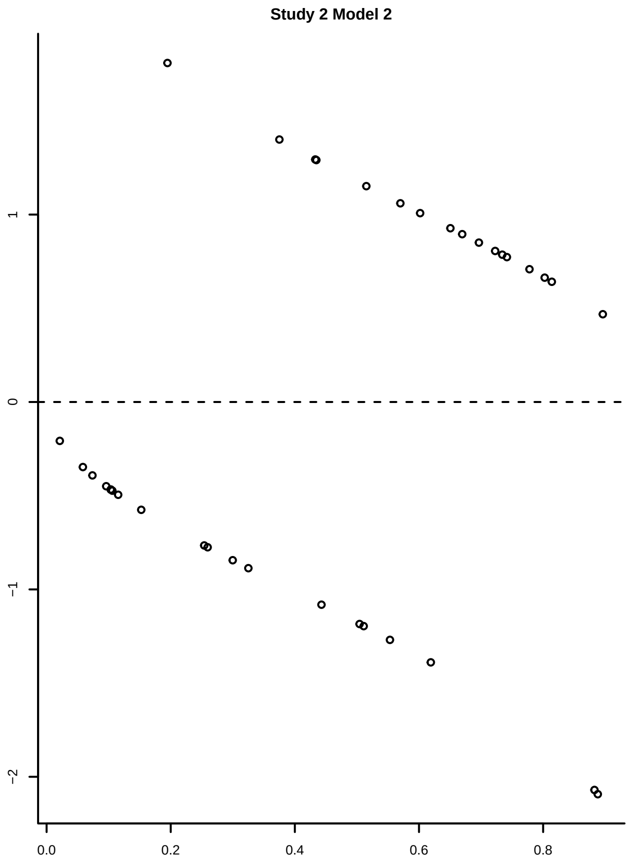
```
s1mod2.fits <- fitted(s1mod2)
s1mod2.dev.resids <- resid(s1mod2)
par(mfrow=c(1,2), cex =0.45, mar = c(4,4,2.3,0.2), bty ="L")
plot(s1mod2.fits, s1mod2.dev.resids, xlab = " Fitted Values ", ylab = " Deviance Residuals", main =  "S
abline(h=0, lty=2)
qqnorm(s1mod2.dev.resids, xlab = "Std, Normal Quantities", ylab = "Deviance Residuals", main = "Study 1
qqline(s1mod2.dev.resids)
```

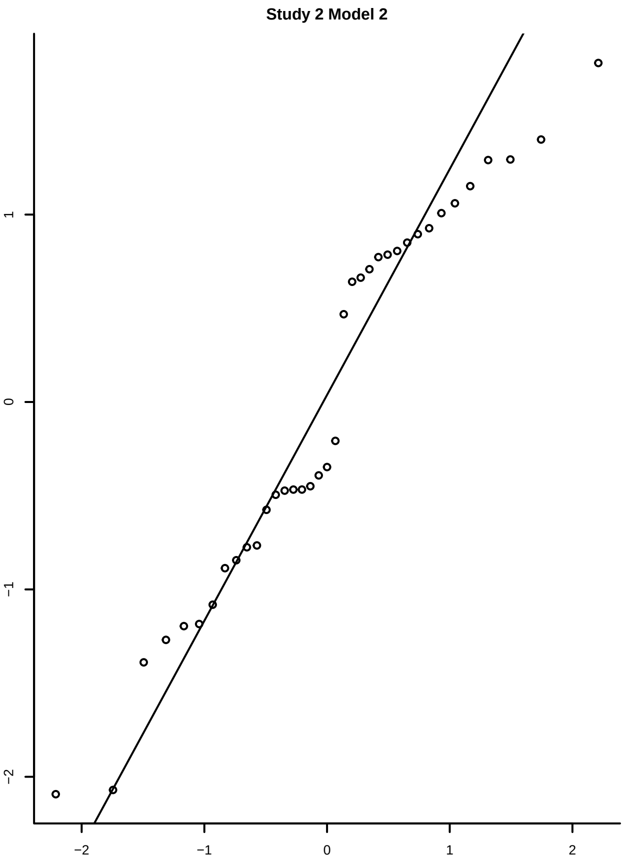


```
# Study 2 Plots
s2mod1.fits <- fitted(s2mod1)
s2mod1.dev.resids <- resid(s2mod1)
par(mfrow=c(1,2), cex =0.45, mar = c(4,4,2.3,0.2), bty ="L")
plot(s2mod1.fits, s2mod1.dev.resids, xlab = " Fitted Values ", ylab = " Deviance Residuals", main =  "S
abline(h=0, lty=2)
qqnorm(s2mod1.dev.resids, xlab = "Std, Normal Quantities", ylab = "Deviance Residuals", main = "Study 2
qqline(s2mod1.dev.resids)
```



```
s2mod2.fits <- fitted(s2mod2)
s2mod2.dev.resids <- resid(s2mod2)
par(mfrow=c(1,2), cex =0.45, mar = c(2,2,2.3,0.2), bty ="L")
plot(s2mod2.fits, s2mod2.dev.resids, xlab = " Fitted Values ", ylab = " Deviance Residuals", main =  "S
abline(h=0, lty=2)
qqnorm(s2mod2.dev.resids, xlab = "Std, Normal Quantities", ylab = "Deviance Residuals", main = "Study 2
qqline(s2mod2.dev.resids)
```





For both studies, a binary logistic model was used. This seemed appropriate since the dependent variable, the sentence outcome, is binary with 0 = life and 1 = death. Our group was able to reproduce almost the exact same results as the study. However, when looking at goodness of fit for the binary logistic regression model it appears the model is not well fitted. When looking at the Deviance vs. Fitted value graphs the data is not randomly scattered around 0, and the deviance residuals do not appear to be normally distributed. This could be due to how it appears that the researchers in the study did not make sure the assumptions of the binary logistic regression model were met before fitting the data. In this study, the probability of getting the life sentence is not constant from inmate to inmate, since that probability depends on the severity of their crime.

The papers final results for study 1 resulted in a 95% CI odds ratio of [0.54,.91] and I think it should have been mentioned that the confidence interval contained .91 which is pretty close to 1 and indicates there is no relationship between trustworthiness and sentencing results. Not to mention the 2nd study's CI for the odds ratio (.05, 1.06) did contain one indicating trustworthiness did not effect sentencing. Considering this, I do not think concluding that untrustworthy faces are more likely to be sentenced to death was appropriate.