Logistic Regression

Daniel Mlinga

Model 1

```r
library(Stat2Data)
data(CreditRisk)
mod1 <- glm(Overdrawn~DaysDrink, data= CreditRisk, family = binomial(link = 'logit'))
summary(mod1)
```

```
Call:
glm(formula = Overdrawn ~ DaysDrink, family = binomial(link = "logit"),
    data = CreditRisk)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.21469    0.19782 -11.195  < 2e-16 ***
DaysDrink    0.04691    0.01642   2.856  0.00429 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 352.55  on 446  degrees of freedom
Residual deviance: 344.83  on 445  degrees of freedom
  (3 observations deleted due to missingness)
AIC: 348.83

Number of Fisher Scoring iterations: 4
```
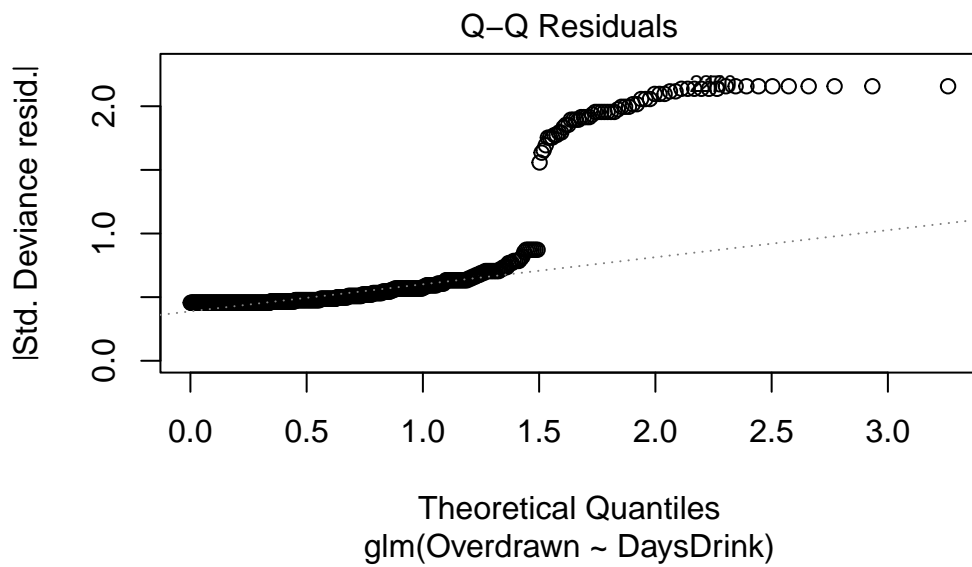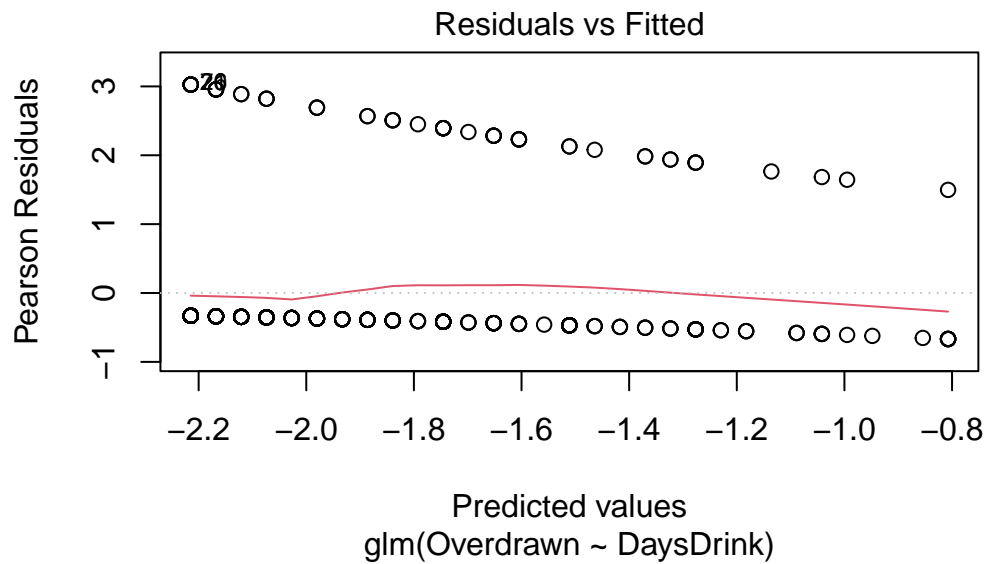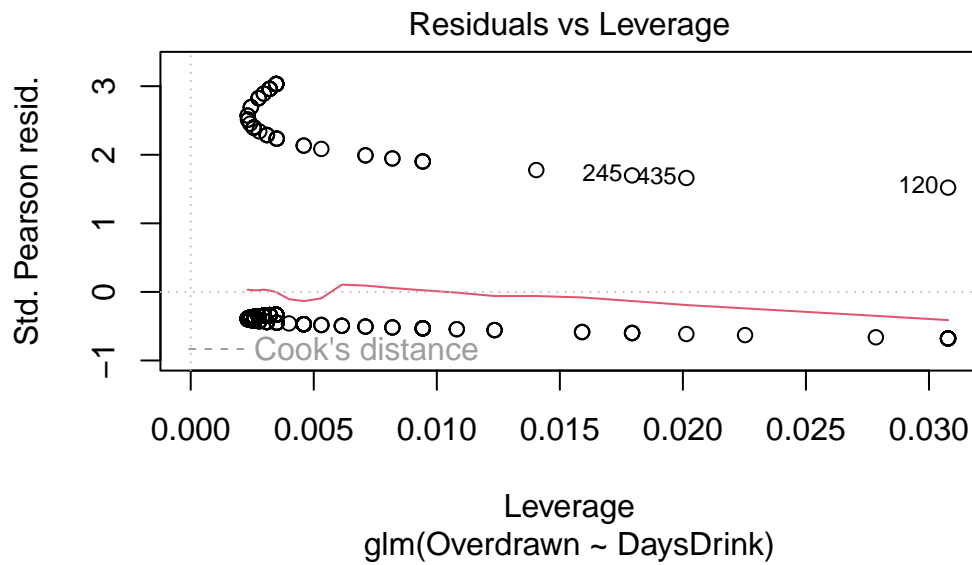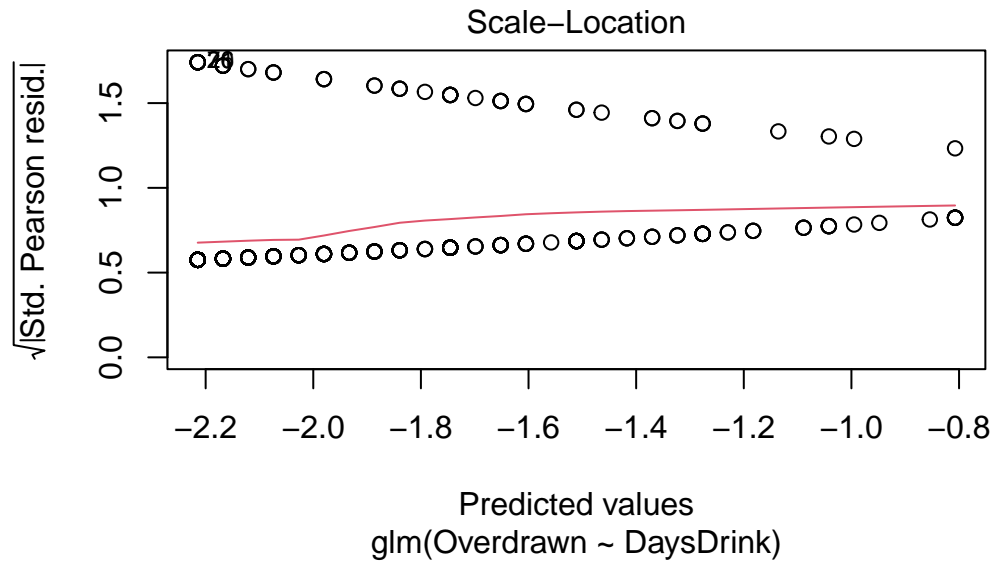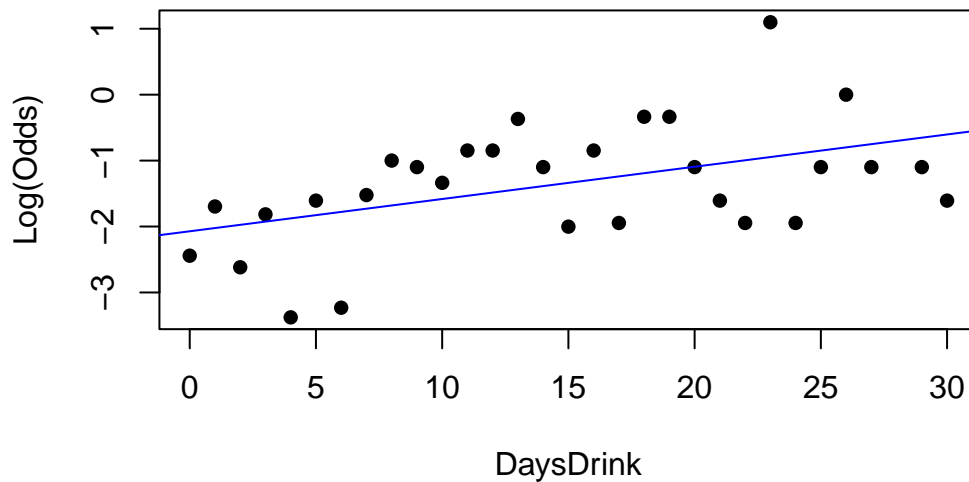
```r
plot(mod1)
```

1

Residuals vs Fitted

Pearson Residuals

Predicted values
glm(Overdrawn ~ DaysDrink)

Q–Q Residuals

|Std. Deviance resid.|

Theoretical Quantiles
glm(Overdrawn ~ DaysDrink)

Scale−Location

√|Std. Pearson resid.|

Predicted values
glm(Overdrawn ~ DaysDrink)

Residuals vs Leverage

Std. Pearson resid.

245 435

120

Cook's distance

Leverage
glm(Overdrawn ~ DaysDrink)

```
emplogitplot1(Overdrawn~DaysDrink, data = CreditRisk, ngroups = 'all', main = "Empirical L
```

Empirical Logit Plot for Model (Overdrawn ~ DaysDrink)


```
xlabel <- expression(Days~Drink)
ylabel <- expression(hat(mu))
```

Model2

```
mod2 <- glm(Overdrawn~Age, data= CreditRisk, family = binomial(link = 'logit'))
summary(mod2)
```

```
Call:
glm(formula = Overdrawn ~ Age, family = binomial(link = "logit"),
    data = CreditRisk)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.73866    1.86146  -3.083  0.00205 **
Age          0.19321    0.09323   2.072  0.03823 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```
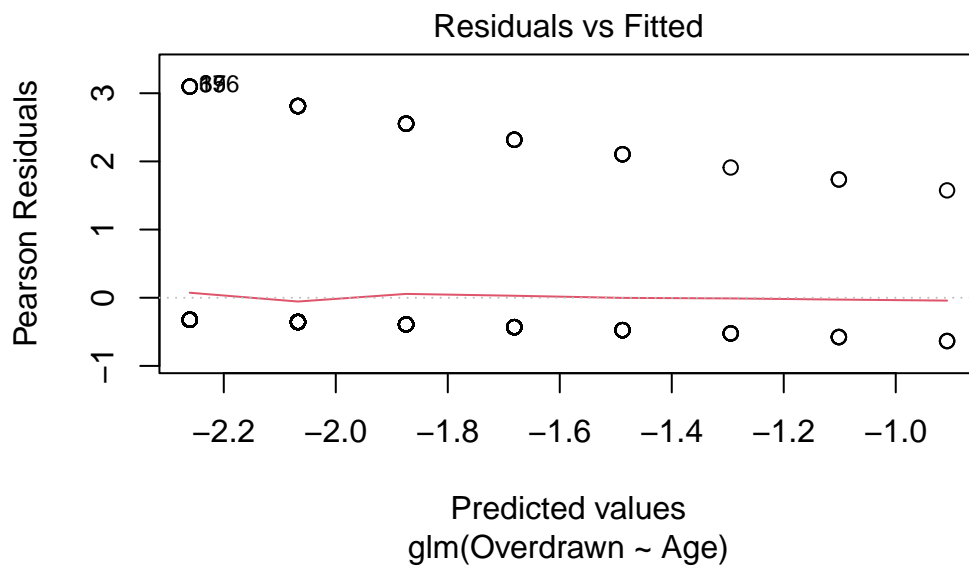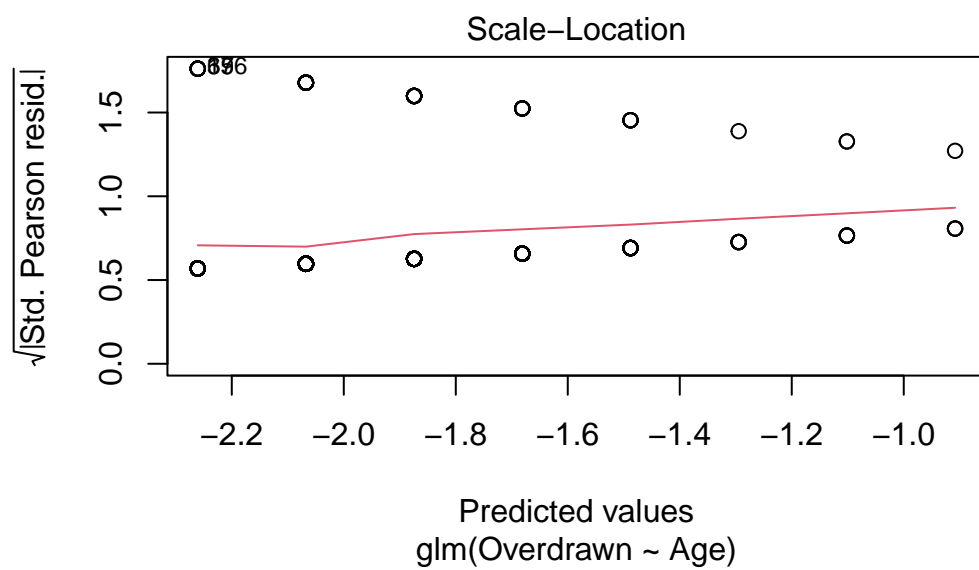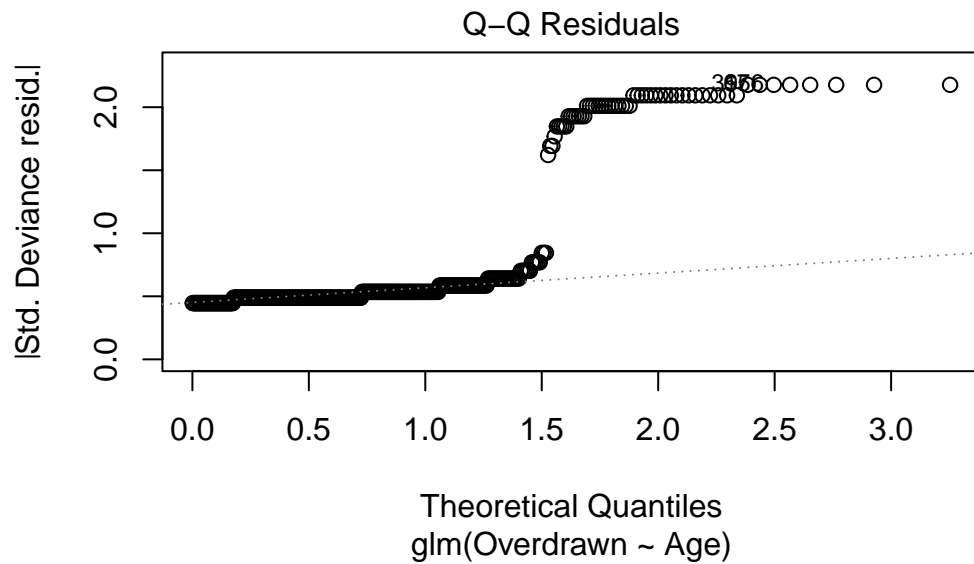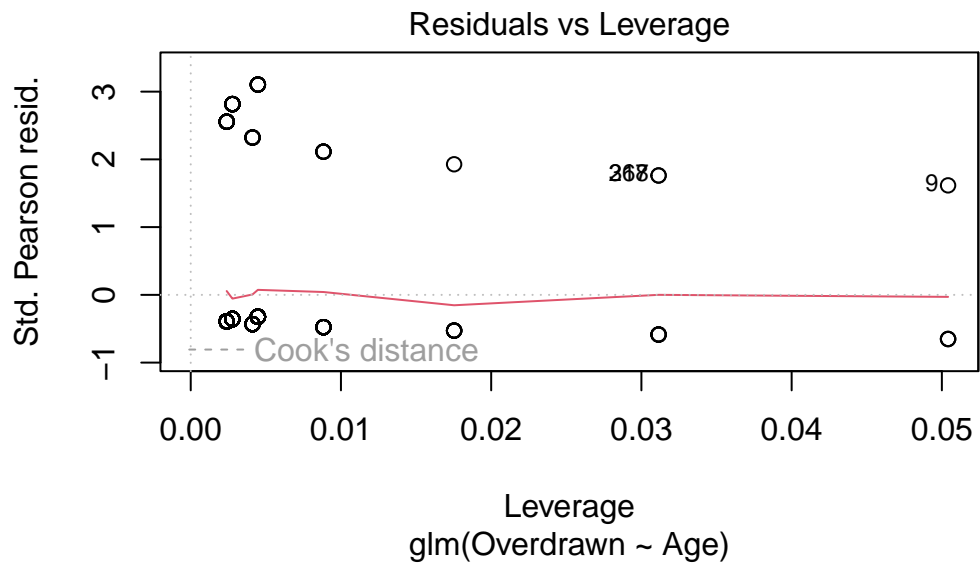
```
    Null deviance: 334.88  on 437  degrees of freedom
Residual deviance: 330.85  on 436  degrees of freedom
  (12 observations deleted due to missingness)
AIC: 334.85

Number of Fisher Scoring iterations: 4
```
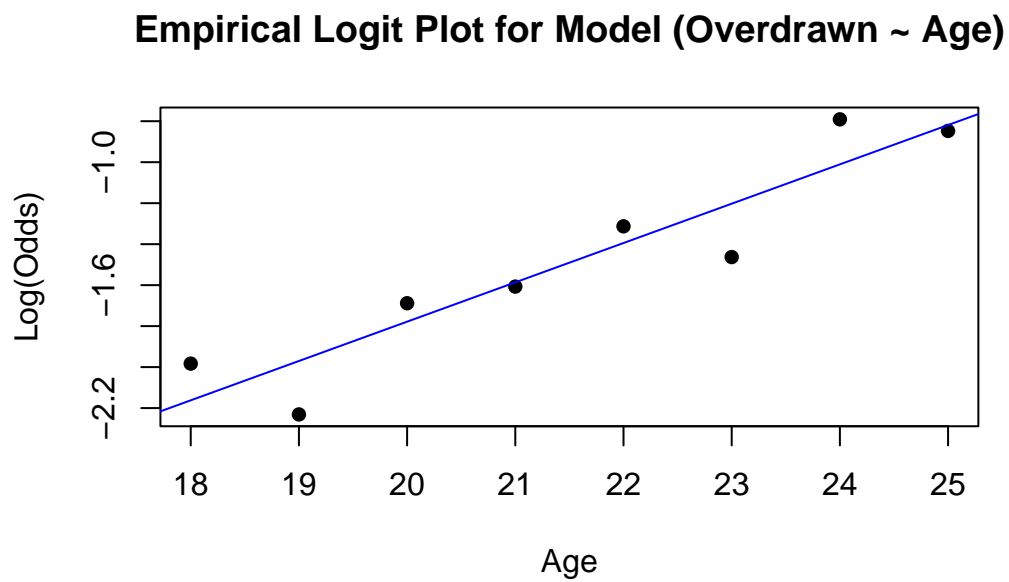
```
plot(mod2)
```

Residuals vs Fitted

Predicted values
glm(Overdrawn ~ Age)

Q−Q Residuals

glm(Overdrawn ~ Age)

Scale−Location

Predicted values
glm(Overdrawn ~ Age)

Residuals vs Leverage

glm(Overdrawn ~ Age)

```
emplogitplot1(Overdrawn~Age, data = CreditRisk, ngroups = 'all', main = "Empirical Logit P
```

Empirical Logit Plot for Model (Overdrawn ~ Age)

```r
  xlabel <- expression(Age)
  ylabel <- expression(hat(mu))
```

Model 3

```r
  mod3 <- glm(Overdrawn~Sex, data= CreditRisk, family = binomial(link = 'logit'))
  summary(mod3)
```

```
Call:
glm(formula = Overdrawn ~ Sex, family = binomial(link = "logit"),
    data = CreditRisk)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.4092     0.2610  -9.231  < 2e-16 ***
Sex           0.8558     0.3093   2.767  0.00565 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 352.26  on 445  degrees of freedom
Residual deviance: 343.90  on 444  degrees of freedom
  (4 observations deleted due to missingness)
AIC: 347.9

Number of Fisher Scoring iterations: 5
```
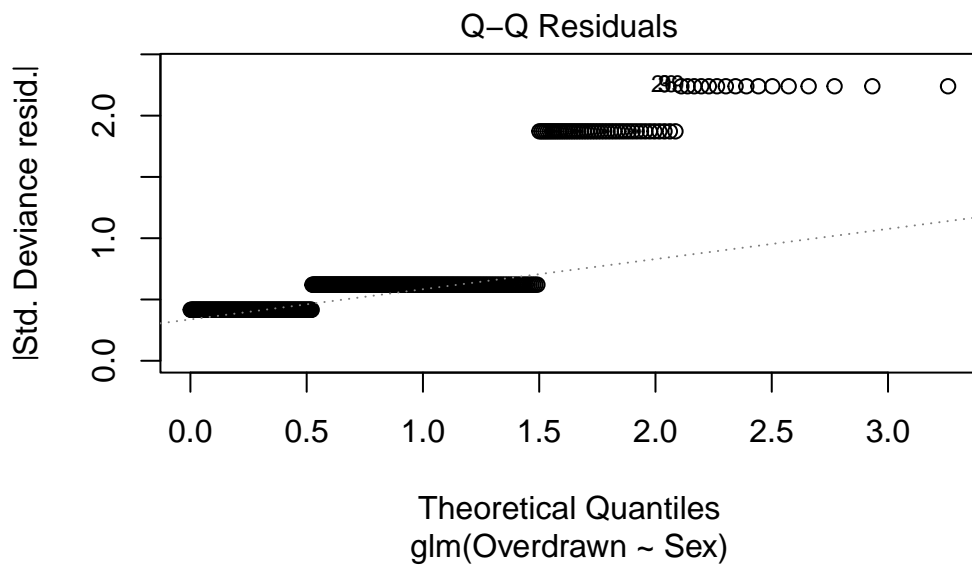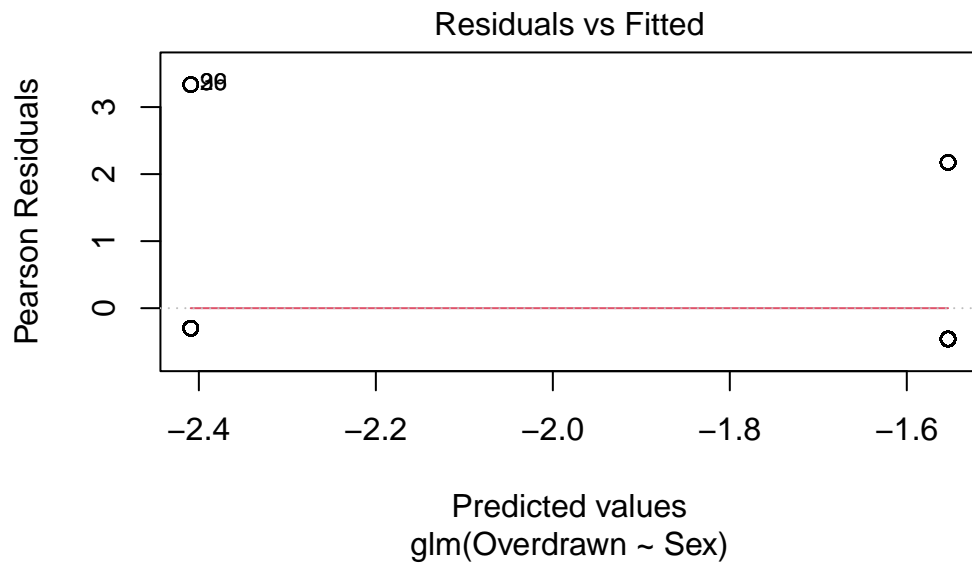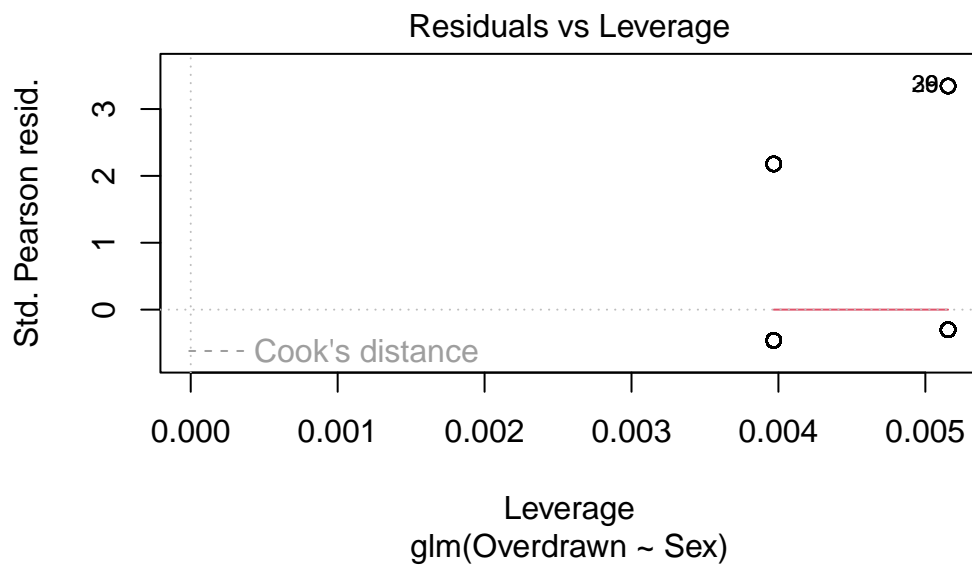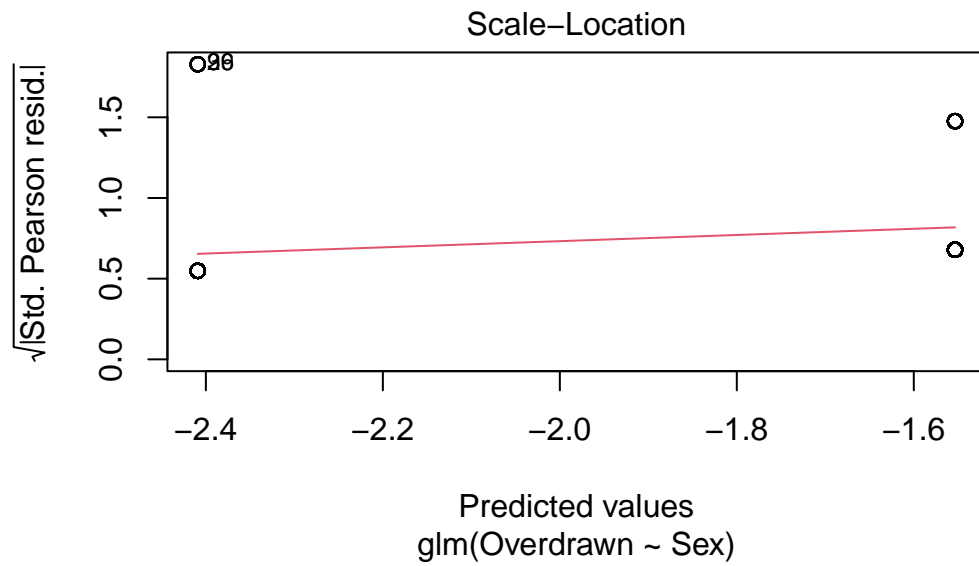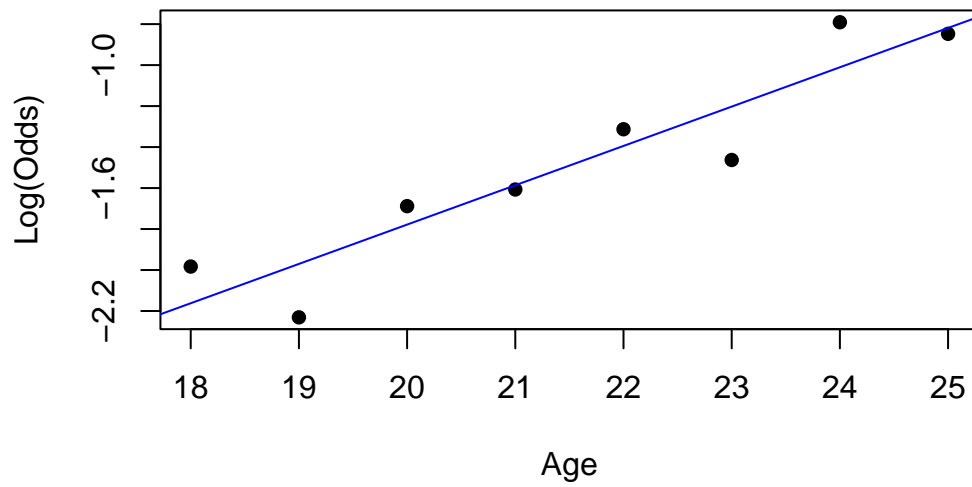
```r
  plot(mod3)
```

Residuals vs Fitted

Predicted values
glm(Overdrawn ~ Sex)

Q–Q Residuals

Theoretical Quantiles
glm(Overdrawn ~ Sex)

9

Scale−Location

Predicted values
glm(Overdrawn ~ Sex)

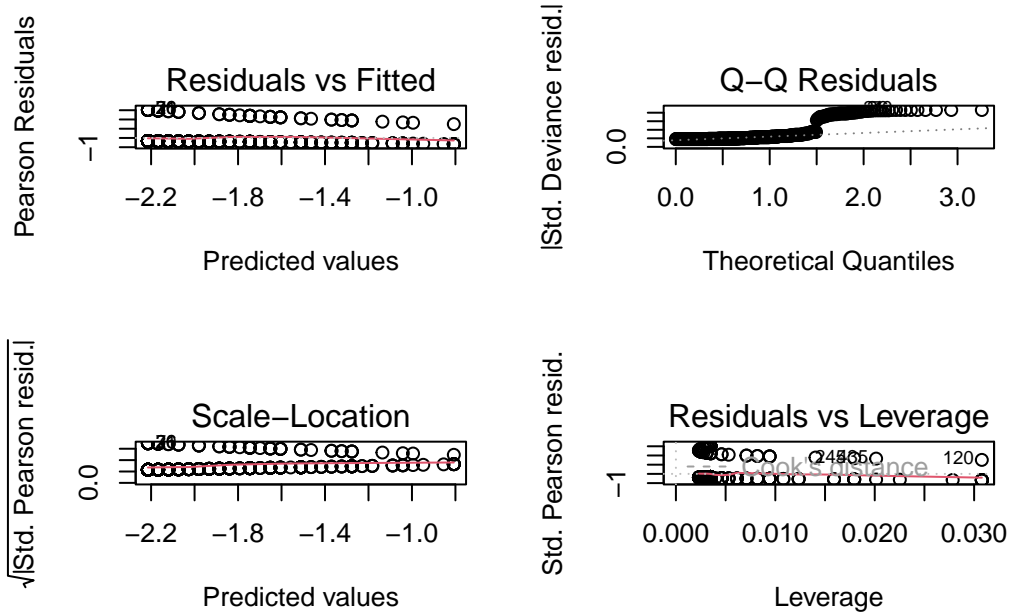## Residuals vs Leverage

Leverage
glm(Overdrawn ~ Sex)

```
emplogitplot1(Overdrawn~Age, data = CreditRisk, ngroups = 'all', main = "Empirical Logit P
```
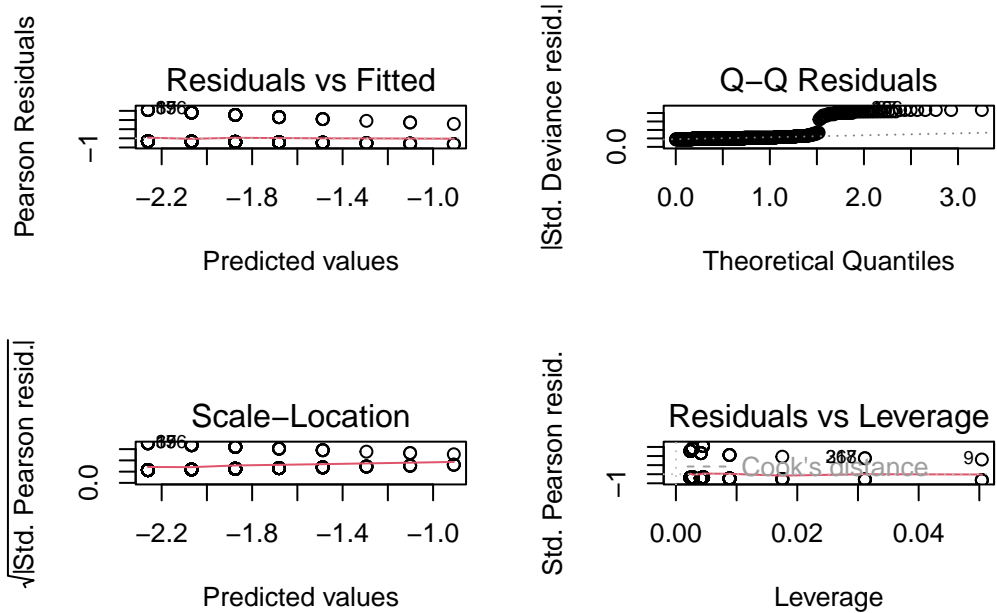
10

Empirical Logit Plot for Model (Overdrawn ~ Sex)



```r
xlabel <- expression(Sex)
ylabel <- expression(hat(mu))


# Diagnostic plots for mod1 (Overdrawn ~ DaysDrink)
par(mfrow = c(2, 2))
plot(mod1)
```

Residuals vs Fitted

Pearson Residuals

Predicted values

Q–Q Residuals

|Std. Deviance resid.|

Theoretical Quantiles

Scale–Location

√|Std. Pearson resid.|

Predicted values

Residuals vs Leverage

Std. Pearson resid.

Leverage

```
# Diagnostic plots for mod2 (Overdrawn ~ Age)
par(mfrow = c(2, 2))
plot(mod2)
```

```
# Diagnostic plots for mod3 (Overdrawn ~ Sex)
par(mfrow = c(2, 2))
plot(mod3)
```
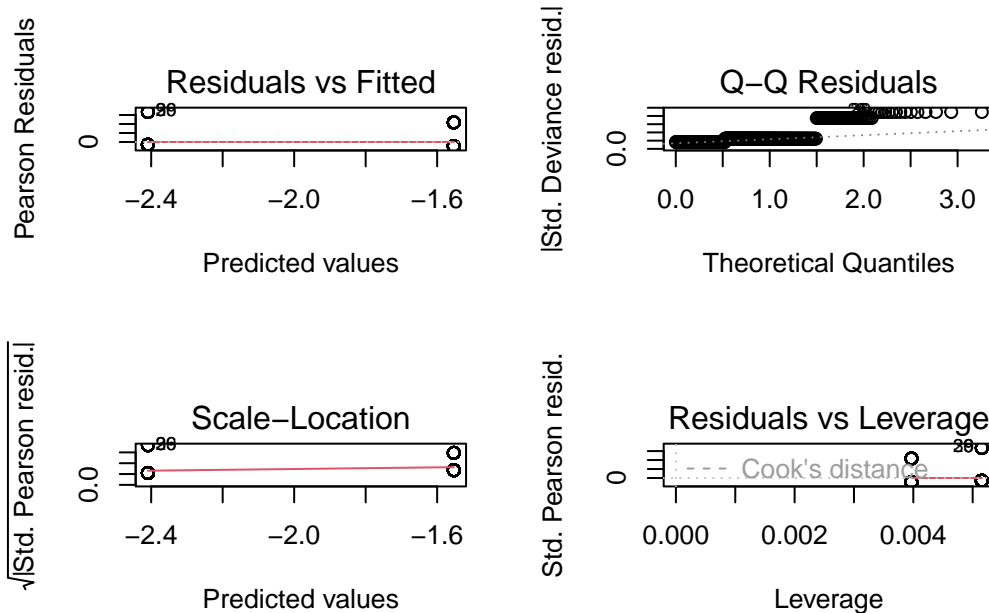
```r
# Extract coefficients and standard errors from the summary output
coef_age <- coef(mod2)["Age"]
se_age <- summary(mod2)$coef["Age", "Std. Error"]

# Calculate odds ratio
odds_ratio_age <- exp(coef_age)

# Calculate confidence interval for odds ratio
lower_ci <- exp(coef_age - 1.96 * se_age)
upper_ci <- exp(coef_age + 1.96 * se_age)

# Print results
cat("Odds Ratio for Age:", odds_ratio_age, "\n")
```

Odds Ratio for Age: 1.213141

```r
cat("95% Confidence Interval for Odds Ratio (Age):", lower_ci, "-", upper_ci, "\n")
```

95% Confidence Interval for Odds Ratio (Age): 1.010536 - 1.456368