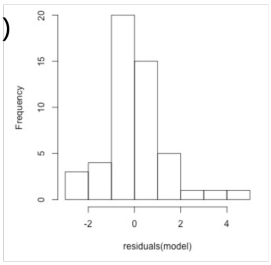
Quiz 3

2023-06-20

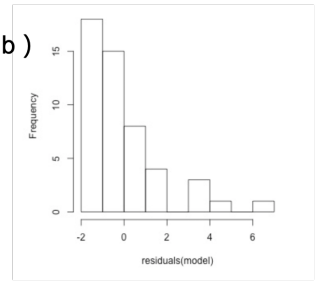
Question 1

Match the following distributions of residuals with the normal probability plots.

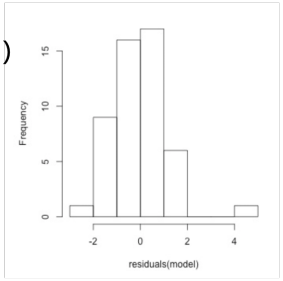
(a)



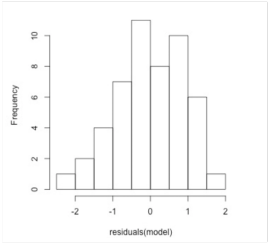
(b)



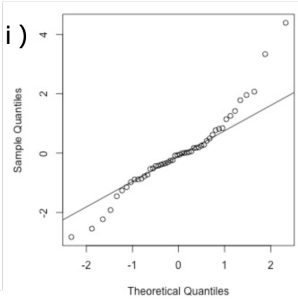
(c)



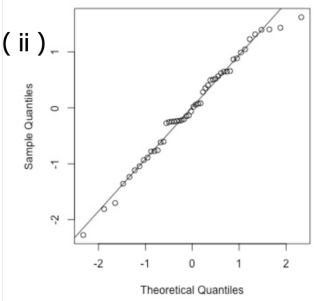
(d)



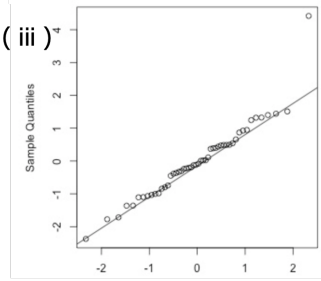
(i)



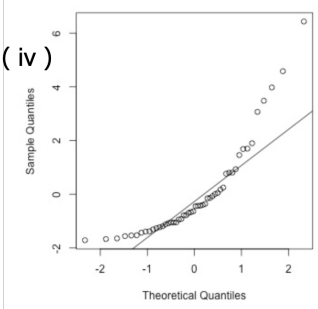
(ii)



(iii)



(iv)



Question 2

In the context of regression analysis, which of the following statements are true?

-I. When the data set includes an influential point, the data set is nonlinear.

-II. Influential points always reduce the coefficient of determination.

-III. All outliers are influential data points.

(A) I only
(B) II only
(C) III only
(D) All of the above
(E) None of the above

Question 3

You are making a regression model for a physical process that you know follows the form $y_i = A + Bx_i + \epsilon_i$. You wish to use regression to find values for A and B. However, you know the variance of the errors is given by $\text{Var}(\epsilon_i) = kx_i^2$ for a positive constant k.

(a) Name the problem with the regression assumptions that you have

(b) You address the problem by transforming your data. You divide both sides of your regression equation by X, to get $Y/X = A/X + B + \epsilon/X$. You then define $Y' = Y/X$, $X' = 1/X$ and $\omega = \epsilon/X$, to arrive at the model equation $Y' = AX' + B + \omega$. State the variance of ω_i, and argue that this transformation has fixed the problem with the regression assumptions

(c) You run the following code in R:

```
xprime <- 1/X
yprime <- Y/X
fit=lm(yprime ~ xprime)
summary(fit)
```

And you get the following output:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.2126     0.1314  117.27   <2e-16 ***
xprime       13.4806     0.7124   18.78   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9446 on 98 degrees of freedom
Multiple R-squared:  0.7826,Adjusted R-squared:  0.7804
F-statistic: 352.8 on 1 and 98 DF,  p-value: < 2.2e-16
```

What are the fitted values $\hat{A}$ and $\hat{B}$? Write a formula for predicting $\hat{y_i}$ from x_i (note: not y' and x'!)

(d) Is there evidence at the $\alpha = 0.05$ level to reject the null hypothesis that A = 15. (The critical value of t(0.025;df=98) = 1.98)