Full Detail

Work out the following problems. **Show your work.**

1. Consider a dataset with 46 observations of patient satisfaction (a score from 0 to 100), and severity of illness (an index from 0 to 100). Part of the output of `summary` of the linear model in `R` is shown below:

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 183.0770    24.3249   7.526 1.95e-09 ***
X1           -2.4093     0.4806  -5.013 9.23e-06 ***
---
Residual standard error: 13.91 on 44 degrees of freedom
Multiple R-squared:  0.3635,Adjusted R-squared:  0.3491
F-statistic: 25.13 on 1 and 44 DF,  p-value: 9.23e-06
```

   (a) Which variable is the response variable and why, and which is the explanatory variable and why?

   (b) Interpret the slope and intercept of the linear regression (if appropriate).

   (c) Test the hypothesis that the explanatory variable has a linear relationship with the response variable, using $\alpha = 0.01$. **State the null and alternative hypotheses, the conclusion, and state and interpret the p-value.**

(d) Assume the range of your explanatory variable is 41 to 62. If we predict two new values of Y at $x^* = 50, 65$, respectively, which of these predictions will have more error? Explain your answer.

(e) Predict the average value of your response variable at the value 53 of your explanatory variable.

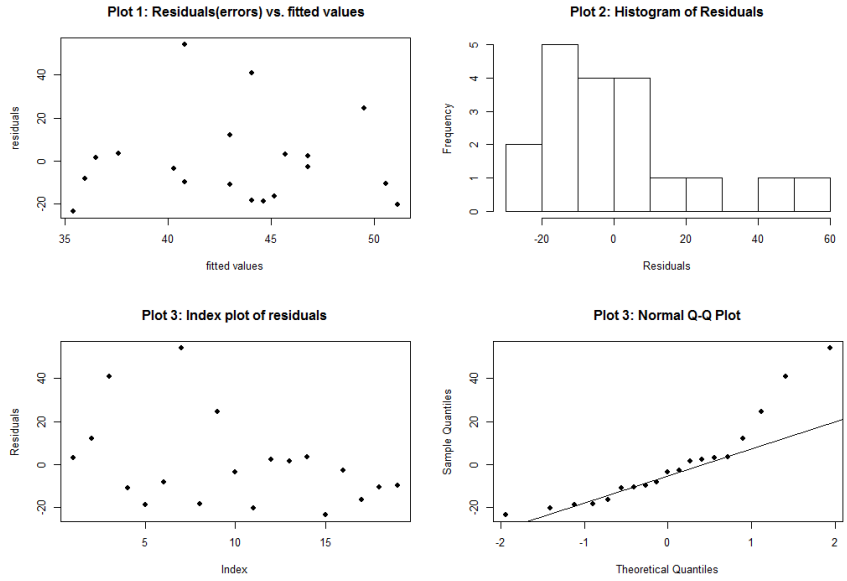
(f) The result of the Fligner-Killeen test is found below:

```
ata:  the.model$residuals and Groups
Fligner-Killeen:med chi-squared = 0.096145, df = 1, p-value = 0.7565
```

State the conclusion in terms of null hypothesis, and interpret the p-value in terms of the null hypothesis.

(g) Interpret the value of R^2 for this linear regression line. Based on this value, do you think the regression model performs well?

2. A new regression has been performed, and this regression has the following diagnostic plots:



(a) List the three of the assumptions of linear regression.

(b) Which assumption does `Plot 1` test? What does this plot suggest for the assumption?

(c) Which assumptions do `Plot 2` and `Plot 4` test? What do these plots suggest for the assumption?

(d) Which assumption does `Plot 3` test? What does this plot suggest for the assumption?

3. Another variable was added to the model - the gender of the patient, with categories `M` and `F`. The estimated coefficients are:

```
Coefficients:
(Intercept)      X1          X2M
184.304       -2.421       -1.250
```

(a) Write down the two separate models for each gender.

(b) Interpret the value of $\hat{\beta}_1$ for the model corresponding to a gender of `M` in terms of the problem.

(c) If a male with a severity of illness score of 50 had a patient satisfaction score of 40, estimate the error in the prediction based on the estimated regression line.

(d) The 95% confidence interval for $\hat{\beta}_2$ is: (-9.634281, 7.135188). Does this suggest that the patients gender had a significant impact on their overall patient satisfaction? Explain.