STA 506 2.0 Linear Regression Analysis

Tutorial 1

A medical research team wants to study the relationship between the **head circumferences of newborns (centimetres) - HC** and the **gestational age (weeks) - GA**. A simple random sample of 57 babies was selected from the records of babies born in a certain hospital, and their head circumferences (centimetres) and gestational age (weeks) were recorded.

In their attempt to find the relationship between head circumferences and gestational age, they plotted the head circumference versus the gestational age and obtained the following graph
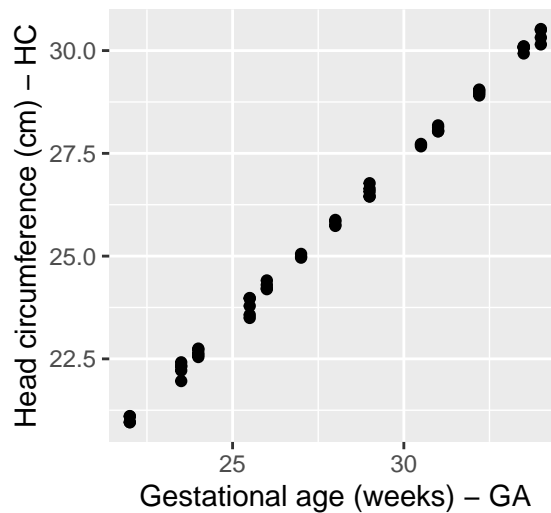

Figure 1: The scatter plot of head circumference versus the gestational age. The Pearson correlation coefficient is 0.99

i) Comment on the scatter plot given above.

ii) Write the model you would fit to these data. Define all terms in it and state any assumptions you make regarding the model.

A simple linear regression analysis was performed with these data and the following outputs were obtained using R software.

Output A

```
Call:
lm(formula = HC ~ GA)

Coefficients:
(Intercept)           GA
     3.9707       0.7781
```

Output B

```
Call:
lm(formula = HC ~ GA)

Residuals:
     Min       1Q   Median       3Q      Max
-0.31535 -0.05153  0.01130  0.07656  0.23882

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.970711   0.113803   34.89   <2e-16 ***
GA          0.778069   0.004003  194.38   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.115 on 55 degrees of freedom
Multiple R-squared:  0.9985,    Adjusted R-squared:  0.9985
F-statistic: 3.779e+04 on 1 and 55 DF,  p-value: < 2.2e-16
```
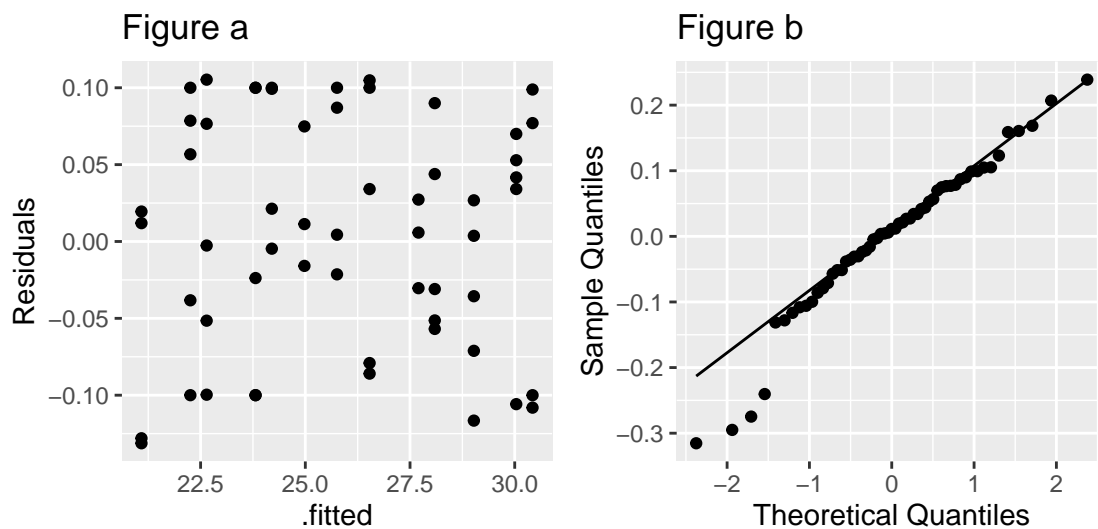
Output C

```
Analysis of Variance Table

Response: HC
          Df Sum Sq Mean Sq F value    Pr(>F)
GA         1 500.02  500.02   37786 < 2.2e-16 ***
Residuals 55   0.73    0.01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output D

Figure a — Figure b

Output E

```
    Shapiro-Wilk normality test

data:  fitmodel$.resid
W = 0.95631, p-value = 0.03828
```

iii) Write the fitted regression model.

iv) Complete the ANOVA table below (Copy the table below in your answer script and complete it.).

Source of variation	DF	Sum of squares (SS)	Mean Square (MS)	F-value	p-value
Regression					
Residual error					
Total					

v) Test the significance of the model fitted. You should clearly write the hypothesis, decision and conclusions.

vi) What proportion of the variation in the response is explained by the model fitted?

vii) Two undergraduates studying Biostatistics were looking at this analysis.

Student A: said that the results strongly suggest that this model is highly significant and can be used for prediction purposes.

Student B: said that the results show the fitted model is not appropriate for this case and this model cannot be used for prediction.

With whom would you agree? Justify your argument using the results given above (Outputs A-E).

viii) The point on the graph (Figure 2, below), coloured in red, represents a baby with gestational age 31 weeks and head circumference 28.03 cm, (GA=31, HC=28.03).

a. Find the fitted value for this point.
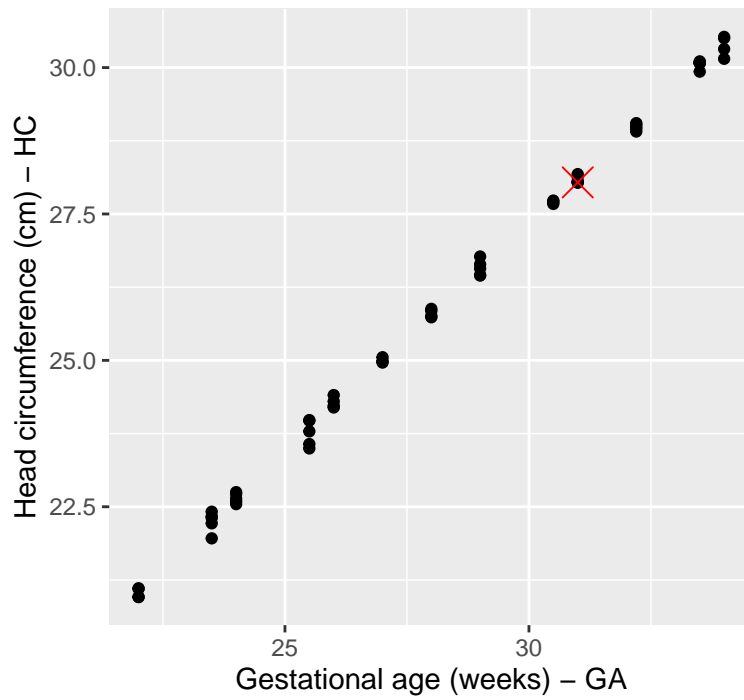
b. Find the residual for this point.

Figure 2: The scatterplot of head circumference versus the gestational age. The point, (GA=31, HC=28.03), is coloured in red on the graph.